

Conceptual frameworks for multimodal social signal processing

Paul M. Brunet · Roddy Cowie · Dirk Heylen ·
Anton Nijholt · Marc Schröder

Received: 24 April 2012 / Accepted: 27 April 2012 / Published online: 26 May 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

This special issue is about a research area which is developing rapidly. Pentland [4] gave it a name which has become widely used, ‘Social Signal Processing’ (SSP for short), and his phrase provides the title of a European project, SSPnet, which has a brief to consolidate the area.¹ The challenge that Pentland highlighted was understanding the nonlinguistic signals that serve as the basis for “subconscious discussions between humans about relationships, resources, risks, and rewards”. He identified it as an area where computational research had made interesting progress, and could usefully make more.

If effective progress is to be made, one of the requirements is to develop some consensus on a variety of issues that are basic to the area—obviously the topics to be covered, but also terminology, the literature that people in the field are expected to know, the simplifications that are considered acceptable, and so on. That kind of statement might look routine, but in the context of technology dealing with

human thoughts and feelings, there is always a grim precedent to consider. Technologies that were supposed to detect lying fell short of any reasonable standards of reliability, and yet they convinced both the public and (for a while) the law [1, 3]. It is not a mistake that should be recycled. More than anything else, that example defines the problem that faces technology moving onto grounds that have traditionally belonged to human judgment: sophisticated technology plus naivete about human beings is a recipe for disaster.

Efforts since Pentland’s paper have made it clear that it is not easy to achieve well-grounded consensus for the new area. Some of the reasons are superficial, such as newcomers assuming that the name defines the field rather than being a useful label for an existing (and expanding) body of work. Others are deeper, such as the fact that there are notoriously intractable divisions in the existing literature on social phenomena (e.g. [2]). Those divisions reflect the uncomfortable reality that social phenomena defy any single, coherent analysis, and it would be naïve to expect that the new field could transcend them. What it can do is to find a way of living with them.

The aim of this special issue is to reflect the kinds of conceptual framework that are emerging in the new field. It accepts that part and parcel of the task is to acknowledge tensions. Because the area is clearly difficult, it takes a twofold approach. The traditional, and much weightier strand, consists of papers that address important parts of the conceptual framework, and that to a greater or lesser extent reflect specific viewpoints. The less conventional strand consists of a statement developed within SSPnet, which forms part of this editorial. It has become known as the Declaration of Belfast.

The papers reflect quite diverse positions. There is perhaps a default position, which is shared (to different extents) by three of the papers. But even within those, there are dif-

¹<http://sspnet.eu/>.

P.M. Brunet · R. Cowie (✉)
Queen’s University Belfast, Belfast, UK
e-mail: r.cowie@qub.ac.uk

P.M. Brunet
e-mail: p.brunet@qub.ac.uk

D. Heylen · A. Nijholt
University of Twente, Twente, The Netherlands

D. Heylen
e-mail: d.k.j.heylen@utwente.nl

A. Nijholt
e-mail: A.Nijholt@utwente.nl

M. Schröder
DFKI GmbH, Saarbrücken, Germany
e-mail: marc.schroeder@dfki.de

ferences of emphasis; and beyond them, there are quite different perspectives to consider.

In the centre, there are striking overlaps between Brunet and Cowie and Scherer et al. Both papers understand the challenge in terms of states to be detected, and signals that carry information about them. The states may be states of an individual, or of a set of people who are interacting, or (Brunet and Cowie say) of an organisation. Both stress that the signals are not neatly packaged: a major part of the challenge is to pull the relevant information from a multimodal flux extended over time.

Scherer et al. develop that general framework in one direction. They offer a succinct list of subject states to be identified (using terms like interested, surprised, stressed, accepting, etc.), and sources of information about them (involving talk style, revealing events, the focus of the speaker, and the dialog role)—these, for them, provide the social signals. They look in some detail at technologies that may be relevant, and develop their ideas through a detailed study of particular data.

Brunet and Cowie move in another direction from the common ground. Their emphasis is on the psychological complexities that have to be reckoned with. They highlight the enormous range of states and signals that may be relevant to social interactions; the different kinds of control that humans may exert over the production of signals, and the different kinds of inference that they may employ; and also the contextual and cultural issues that bear on the generation & interpretation of social signals. They do not argue that systems should reproduce the complexities of human processing, but that system developers should be alert to it.

In both Scherer et al. and Brunet and Cowie, the states most often discussed are socially significant states of an individual. However, Brunet and Cowie acknowledge in principle that some significant states are intrinsically concerned with relationships between interactants. Janssen looks in depth at a key kind of relational state, which is empathy. He argues that in fact, empathy needs to be analysed on different levels—not only cognitive empathy, which has dominated previous research, but also emotional convergence and empathic responding. That emphasis means that processing has to be concerned with relationships between signals recorded from different individuals, which in turn raises challenges for data capture and analysis.

A sharply different approach appears in the paper by D'Errico, Poggi and Vincenze. In the papers considered so far, non-verbal signals are generally thought of as conveying information which is qualitatively different from most of the information conveyed by verbal signals—broadly speaking, about global states of speakers and their relationships. D'Errico et al. reflect a tradition which considers non-verbal communication as 'body language' in a very literal sense, consisting of communicative acts whose meaning could be

expressed in words, but happens not to be [5]. For example, slow headshakes are taken to convey "I can't believe that he is so hopelessly stupid". The more general categories onto which they map non-verbal behaviours are drawn from speech act theory and rhetoric rather than psychology. They use studies of political discourse to show how various moves can be used in combination with language to discredit opponents, and introduce a system for coding them.

Mehu et al. address the issue of divergence itself rather than presenting a position of their own. They emphasise that the divergence has roots in the material on which the emerging discipline has to draw, and they stress the need for a sophisticated attitude to that material. They address the issue at two main levels, vocabulary and overarching concepts. At the level of vocabulary, they set out an extensive list of key terms, and describe the different meanings that the terms carry in different disciplines. At the level of overarching concepts, they discuss different conceptions of information and meaning in general, and then of social signals in particular. They advocate a pluralistic response, and regard it as "the responsibility of each SSP scholar to get familiar with the different approaches".

It is right and proper that the papers in the special issue should reflect different approaches. However, it is also important find ways of defining common ground. That is what SSPnet set out to do in the Declaration of Belfast, which is included here by permission of the SSPnet project members.

1 Declaration of Belfast

Social Signal Processing (often abbreviated to SSP) is an emerging field. The aim of this declaration is to express the way the field is understood by people who are currently active in it. They have come into the field from diverse discipline backgrounds, and are members of the SSPnet Network of Excellence. It is normal that the exact boundaries of a field become clearer as research progresses, and SSP can be expected to follow the same pattern.

2 Brief statement

Social Signal Processing studies 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.

It is a collaboration between research traditions in technology and human sciences, increasingly developing an interdisciplinary identity.

3 Key goals of SSP research

The goals of SSP research can be classified under three headings: technological goals, human science goals, and practical impact goals.

3.1 Technological goals

- (1) To develop systems capable of detecting and interpreting behavioural patterns that carry information about human social activity (analysis).
- (2) To develop systems capable of synthesising behavioural patterns that carry socially significant information to humans (synthesis).
- (3) To develop systems capable of using patterns that carry socially significant information to synthesise appropriate behaviours in an interaction (responsiveness).

3.2 Human science goals

- (1) To develop theories regarding the use of social signals during human-human interactions that can inform artificial agent behaviour, and can inform human-computer interactions.
- (2) To contribute to the human science literature by modifying current theories and proposing new theories informed by the computational research in SSP.
- (3) To create databases suitable for the analysis of human-human interactions, and suitable for training synthesis systems.
- (4) To develop representational systems that describe human social behaviour and cognition in ways that are appropriate to technological tasks (such as labelling databases).
- (5) To develop methods of measuring & evaluating social interactions (human/human and human/machine).
- (6) To develop sophisticated tools for instrumenting human science research.

3.3 Practical goals

Application of the research is not restricted to a narrowly predefined set of issues. It aims to address practical problems in a range of areas. Application has already begun in some areas, and others can easily be foreseen. Natural application areas include

- Artificial agents (e.g for advertising, customer services)
- Ambient intelligence

- Artificial companions
- Assisted living
- Entertainment
- Education
- Human-computer interactions
- Monitoring in health care
- Social skills training
- Multimedia indexing.

4 Key topics

Research in Social Signal Processing recognises the significance of a wide range of topics that have been studied in the human sciences. Some of these define topics that are likely to be the focus of particular projects in SSP; others are overarching in the sense that they affect most SSP research. Many of them are reflected in the thematic work packages in the SSP Network. The following list identifies some of the key topics.

- The range of relevant signals
- The ways in which signals interact and combine in real interactions
- The ways in which signals depend on culture & social identity, and carry information about them
- The ways in which signals depend on power relations, and carry information about them
- The ways in which signals indicate deception & authenticity
- The ways in which signals contribute to influence, credibility & persuasiveness
- The role of context in the production and interpretation of social signals
- The relationship between voluntary and involuntary signalling
- The relationship between awareness of social signals and response to them
- The nature of social meaning.

5 Key challenges

The domain of SSP has specific challenges arising from the nature of the research and to the strong collaboration between human sciences and technology research. The challenges are not only achievable, but should be considered paramount to the success of SSP research and the SSP Network. A list of the core challenges is provided.

- To develop suitable database resources
- To match existing databases with available technologies, i.e.
 - to develop technologies that can work with existing (and conceivable) databases

- to develop databases that can work with existing (and conceivable) technologies
- To collect knowledge about the patterns of signals to be analysed and synthesised that is at an appropriate level of detail to inform SSP technologies
 - existing literatures often do not approach the necessary level of detail
- To develop models of individuality (e.g. personality, culture, identity, stance) that are suited to computational work
- To develop models of impression formation that are suited to computational work
- To develop methods of modelling behavioural dynamics
- To develop analyses that capture causal relationships
- To develop suitable ‘mid-level’ perception techniques (e.g. constancy, segregation)
- To develop controllable, high-quality synthesis techniques.

6 Emerging balances

Some issues with a significant bearing on the character of the field are still a matter of debate. Although they have not been decisively resolved, the profile of activity in SSPnet implies that it tilts towards a particular kind of balance. Key examples are the following.

- Is language included? From a human science standpoint, language is the social signal *par excellence*, and should obviously be included. Technologically, there is an obvious motive not to emphasise it: the natural medium of language, fluent, idiomatic speech, is very difficult to handle. The balance implicit in SSPnet is that language needs to be addressed, using transcripts if necessary: however, it is legitimate to give special attention to tasks where the limitations of language processing are not critical.
- How should naturalness and artificiality be balanced? Research in some related areas has relied heavily on data from actors or laboratory tasks, because naturalistic data is too difficult to find or to analyse. In return, some critics imply that only research on totally natural data is of any value. The balance implicit in SSPnet is that naturalness is a matter of degree. Simulation is acceptable, and probably practically necessary, so long as the signs in question are actually being used in an appropriate kind of interaction.
- What are the appropriate criteria of validity? Research in some traditions insists that data should be associated with a clear ground truth. In SSP that leads to very difficult demands—asking, for instance, what a person *really* felt or intended in a particular situation. A common alternative is to require high inter-rater agreement. That, too, is problematic, because it is a feature of some social signals that different people ‘read’ them in different ways. The

balance implicit in SSPnet is that the appropriate test depends on the application.

7 Interactions between SSP and other disciplines

It is an integral part of establishing SSP to establish appropriate relationships with related disciplines.

One key issue is recognising how much SSP stands to gain from older disciplines. Resources that it can assimilate include not only knowledge (see above), but also techniques (e.g. labelling, experimental designs, standard measures), representational devices (e.g. markup languages), and technical vocabulary.

The interaction between SSP and these disciplines should not be one-sided. SSP research could and should also contribute to other disciplines and help to inform them. The interdisciplinary nature of SSP research provides an incentive to explore ways of integrating material from different disciplines. Attempts to implement ideas also classically contribute to understanding their limitations. SSP also offers disciplines that can be seen as esoteric new kinds of practical application.

The interaction also needs to acknowledge academic realities. The discipline will not retain active input from specialists in a related discipline unless they are able to publish articles that are recognised as contributions to their home discipline.

8 Ethical obligations

SSP deals with issues that are ethically sensitive. As a result, it has a range of ethical obligations. Many are standard, but some are not.

Obligations that are shared with many other fields include

- avoiding distress, deception and other undesirable effects on participants in studies
- maintaining the confidentiality and anonymity of participants involved in the research
- avoiding the development of systems that could reasonably be regarded as intrusive
- limiting opportunities for abuse of the systems that they develop (probably through licensing arrangements)

Particular obligations arise from the combination of complexity and sensitivity that is associated with social signals. The general requirement is sensitivity to the ways that social communication can affect people. Applying that to specific cases depends on intellectual awareness

- of individual issues (personality, age, etc.)
- of cultural issues (norms, specific signs, etc.)

- of general expectations (what is disturbing, humiliating, etc.)

Communicating about the area to non-experts raises particular issues. People are prone to systematic misunderstanding of SSP-type systems, so that they rely on them when they ought not to, fear them when they have no need to, and so on. Obligations relevant to offsetting that are

- honesty, i.e. ensuring that what is said about a system is true;
- modesty, i.e. taking pains to ensure that its limitations as well as its achievements are understood;
- public education, i.e. trying to equip people with the background knowledge to grasp what a particular system might or might not be able to do.

9 Conclusion

It does not seem to be in doubt that there will be a deepening engagement between computing and spontaneous, multimodal communication between humans. The challenge is to ensure that the development avoids some of the pitfalls that are commonplace when technological development is guided by preconceptions about the humans that will use and interact with it, rather than by an empirically grounded understanding of the complexities and subtleties that are actually characteristic of human nature and social processes. The papers in this special issue present resources that can be used to meet that challenge.

It is not to be expected that they will close the subject. On the contrary, one of the most useful outcomes that the issue could generate is debate informed by awareness of the different perspectives that are relevant to it. It would be quite a remarkable achievement if a multidisciplinary area could achieve that level of maturity within a few years of its emergence.

Acknowledgements Work on this editorial and special issue was supported by Work on this article was supported by the European Network of Excellence SSPNet (grant agreement No. 231287).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. APA (The American Psychological Association) (2004) The truth about lie detectors. Downloaded 18.4.2012 from <http://www.apa.org/research/action/polygraph.aspx>
2. Haslam SA, Parkinson B (2005) Pulling together or pulling apart? Towards organic pluralism in social psychology. *The Psychologist* 18(9):50–554
3. Lykken D (1998) *A tremor in the blood: uses and abuses of the lie detector*, 2nd edn. Perseus, New York
4. Pentland A (2007) Social signal processing. *IEEE Signal Process Mag* 24(4):108–111
5. Poggi I (2007) *Mind, hands, face and body. A goal and belief view of multimodal communication*. Weidler, Berlin