Social Signal Processing: What are the relevant variables? And in what ways do they relate?


Document Version:
Early version, also known as pre-print

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Abstract

Studies of the processing of social signals and behaviour tend to focus intuitively on a few variables, without a framework to guide selection. Here, we attempt to provide a broad overview of the relevant variables, describing both signs and what they signify. Those are matched by systematic consideration of how the variables relate. Variables interact not only on an intrapersonal level but also on an interpersonal level. It is also recognised explicitly that a comprehensive framework needs to embrace the role of context and individual differences in personality and culture.

1. Introduction

A landmark review of psychology concluded that, after a century of research, it had reached the stage of trying systematically to identify the relevant variables [1]. It was not a nihilistic conclusion. The point was that a great deal of effort had been expended discovering that the relevant variables were not obvious.

Social Signal Processing should not need to spend a century reaching the same stage, because several disciplines – psychology being one – have worked to identify relevant variables. However, because the literature is large and diverse, it is easy to drift unintentionally into assuming that the variables are obvious. The aim of this paper is to offer the emerging computational discipline of Social Signal Processing a structured overview which helps to offset that tendency, and highlights some potentially relevant variables.

It is obviously not possible to review all that is known about social signal processing in a short paper. But it seems possible to provide a broad layout of the relevant literatures, and that is what the paper aims to do.

2. Alternative frameworks & models

Research from various disciplines has focused on determining how humans detect, interpret, and classify social signals, and consequently how this information affects behaviour during social interactions, and has proposed models and frameworks to explain and represent the processes involved in human-human interactions. These models are discipline-specific in the sense that they are produced by a methodology that lends itself to exploring some of the relevant variables and relationships involved in human social behaviour, but not all. Hence they tend to highlight one or two aspects of social behaviour and social cognition, and gloss over other relevant aspects. A key challenge for Social Signal Processing is to incorporate their different strengths into a comprehensive framework.

Experimental psychologists have worked extensively on the non-verbal signals that humans display and perceive during interactions, with particular emphasis on the face and the emotions that it conveys. Seminal work by Paul Ekman proposed 6 basic emotions (i.e. happiness, sadness, anger, fear, disgust, and surprise), each of which has a corresponding and universal facial expression [2, 3]. That has generated a large body of work on how humans can accurately detect emotional states by picking up on facial signals commonly associated with the basic emotions (e.g. a dropped jaw with relaxed lips and raised eyebrows associated with surprise [4, 5]. From there, research has diverged into topics from the ability of individuals with psychological disorders (e.g. autism [6], schizophrenia [7]) to identify emotional expressions, to sex differences [8], to developmental pathways [9]. Methods that promise objective measurement, from physiology to eye-tracking, have been eagerly embraced.

Linguists and psycholinguists, on the other hand, have focused on issues such as the structure implicit in dialogue. Some ideas, such as the pragmatic analyses associated with Grice and his colleagues, have become very well known. Others, particularly the analyses that deal explicitly with exchanges between two parties, are less familiar. For example, it is widely accepted that conversations occur on two tracks [10]. The first and main track represents the dialogue dedicated to the clarification and grounding of the main track’s information. The second track represents the dialogue dedicated to the clarification and grounding of the main track’s information.

Sociolinguists, in contrast, have focussed on the way speech encodes information about the social affiliations and aspirations of speakers, and relationships between
and determining sarcasm or sincerity, nervousness or instance, intonation plays an essential role in projecting information not provided in the verbal content [14]. For many features of the way people speak are carry disclosure, question, and request) [13, 14]. Additionally, vocabulary, and purpose of the statements (e.g. self-disclosure, question, and request) [13, 14]. Features such as dialect and lexical selection play important, and quite complex roles here.

Anthropologists, like linguists, rely heavily on verbal data when developing models. They have proposed that there are universals in language that influence the foundation of social interactions. For example, an assumption of Brown and Levinson’s model of politeness [11] is that all individual enter social interactions with the mutual understanding that both parties want to protect their ‘face’ and avoid harming the other person’s ‘face’. Face, the individual’s public self-image which they do not want compromised by humiliation, has a positive and a negative subtype. Positive face represents the desire for approval by others, whereas negative face is the desire for freedom of action. All adult individuals know that they and everyone else has these desires.

One of the factors that distinguishes different approaches is that they are associated with different applications. From that point of view, Social Signal Processing is a discipline whose applications are very different from, for instance, language teaching, or psychological therapy. Hence it is appropriate that it should attempt to develop its own framework.

3. Towards a comprehensive framework

To achieve a framework that suits Social Signal Processing, a new model should strive to incorporate lessons from all the approaches outlined above. That cannot be done simply by adding together ideas from the various disciplines: the result would be an amorphous mass. To avoid that, a framework is needed that is capable of giving the relevant pieces a meaningful place. A first step towards that is to enumerate all the relevant elements – the potentially relevant signals, and the things they may signify; and to consider the ways in which the elements may relate.

3.1. Broad categories of variables

A social signal, as defined by Grammer et al., [12] is a messenger carrying information between a sender and a recipient. The sender encodes information into the signal, which is then detected and decoded by the receiver. The signal is transferred over a communication channel which can take many forms. A description of the possible forms of signals follows.

3.1.1 Verbal characteristics.

Spoken language needs to be part of a comprehensive framework to understand social interactions. Not only is the content of the exchange important, individuals can also gather information based on the sentence structure, vocabulary, and purpose of the statements (e.g. self-disclosure, question, and request) [13, 14]. Additionally, many features of the way people speak are carry information not provided in the verbal content [14]. For instance, intonation plays an essential role in projecting and determining sarcasm or sincerity, nervousness or confidence, and approval or disgust, to name a few. The absence of that information has made computer-mediated communication more likely to lead to misunderstandings. When emailing and instant messaging first became popular, individuals did not know how to properly project or interpret tone (e.g. sarcasm) from the written exchange [16]. To help with that, emoticons and net lingo (e.g. writing ‘lol’ to signify ‘laugh out loud’ to project humorous tone) have arisen to inject tone into written exchanges.

3.1.2 Facial characteristics.

During social interaction, people extract a substantial amount of information about emotional states from facial expression rather than from verbal content [17, 18]. Recent research emphasises that in natural data, the signs are not simply archetypal ‘snapshots’: they are distributed over time [19] and linked to body movements [20].

It is also important not to focus exclusively on the facial surface. In particular, the eyes have been identified as a salient feature for conveying not only emotional states, but also determining intent [21]. Some functions are inherently interpersonal, such as establishing joint attention via eye gaze [22]. Joint attention refers to when two individuals are focused on the same event, object, or person. For example, if John makes a flattering comment about a third party who is standing near him to his friend Bill. Bill can identify who John is talking about by tracking John’s eye gaze, and can then provide his own opinion. John does not have to state who he is talking about, his eyes gives that information. Channels of that kind are complex, but would pose real difficulties for an artificial system that could not use them.

3.1.3 Body characteristics.

In addition to the face, the body provides other relevant signals. Body posture (e.g. standing straight up, being slouched over, arms crossed) are useful social signals that have inspired the idea of body language [23]. Physical gestures also contribute to the information conveyed by the body. Many gestures can easily be identified and interpreted during social interactions (e.g. hand waving to say hello), which can indicate politeness, friendliness, aggression, and so on. Furthermore, the physical distance between both individuals can signal intent. For example, if a woman stands in very close proximity to a man, it could indicate attraction. If the woman keeps a noticeable distance stands in very close proximity to a man, it could indicate interest in his advances [24].

3.1.4 Physiological characteristics.

Physiological reactions can provide useful social and emotional information [25]. Some physiological reactions are undetectable without the aid of machines. For example, detecting someone’s frontal EEG patterns is not possible unless they have electrodes on their head.
However, other physiological reactions are not only noticeable, but important social signals. For example, blushing, blinking rate, and sweating are signals that are detectable and may indicate being nervous, aroused, or embarrassed.

3.1.5 Other physical characteristics.
Other physical characteristics can influence social behaviour and signals, and provide useful information. The most salient of these observable physical characteristics related to gender. For example, short hair, a beard, an Adam’s apple, and a flat chest are clear indications that the person is a man. These characteristics will influence the interpretation and the production of social behaviour. For example, a man is more likely to make emotional self-disclosures to a woman rather than another man [26].

Height and especially weight [27] are also social signals that can convey, but not always accurately so, useful information about the other person. A man who is physically fit is more likely to be asked questions about how to get involved with the local sports team than a man who is overweight.

Signals relating to age (apparent or chronological) also come into play. Apparent age is a social signal that the other person may detect and consequently affect their behaviour. Chronological age will influence the person’s own behaviour and interpretation of the other person’s signals. A 60 year old man and a 20 year old man may have different criteria for determining what signals of politeness.

3.1.6 High order characteristics
Information is also carried by what can be described as high order characteristics. These do not represent distinct variables, but instead refer to the quality of the other variable sets.

One such quality is the intensity of the signals and behaviour. The velocity and magnitude of a gesture, or the intensity of a facial expression, can be fundamental to the information that it provides.

Voluntary control over the signals can also be indicative of the person’s intentions. A simple example is that of a smile. A child asks her father if he likes her art work. Smiling automatically would suggest that he genuinely liked the art; a forced smile could be read as concealing negative less positive opinion.

Patterning over time takes a wide range of forms. As a relatively simple example, recognising that an individual always speaks quickly or blushes easily and frequently affects their significance as social signals. At a much higher level, the pattern of interchanges between individuals in a group may indicate which has the role of chairperson, or acknowledged leader.

3.2. What variables may signify
Identifying the variables is one of two parallel tasks. The second is to establish what the variables may signify.

The most obvious is that the signals or behaviour are providing an explicit message that one person is trying to convey to the other. That is the natural way to think about linguistic communication, and it transfers to many kinds of non-linguistic communication too. However, that is only one possible kind of significance.

A second kind of communicative significance involves implicit meanings which the person may not consciously be trying to project. A clear example is that signs and behaviours can indicate the person’s cognitive and affective states. For example, facial expressions are usually presumed to signify affective states. A smiling face is supposed to represent a happy affective state. Note, though, that there are other interpretations – a smile may be more analogous to a speech act.

A related set of possibilities, with strong links to philosophy and AI, is that variables may serve to convey a person’s beliefs, desires, intentions, and attitudes. These may be conveyed in part or in whole by language, but many systems can contribute.

Various kinds of social position and relationship may also be signalled. Gender, age and fitness were mentioned in section 3.1.5. A wide variety of signals may be used to project dominance, authority, respect, or affinity. People’s impressions of these attributes have a major impact on the form and success of social interactions with others.

More abstract analyses have been developed, and have a great deal to offer the field. There have been sophisticated attempts to operationalise the nature of a goal of a communication, invoking both proximal and distal explanations. Theories in ethology imply that there tend to be strong evolutionary pressures towards manipulation in social signals [28]. In contrast in theories concerning the evolution of human language there have been attempts to explain how cooperative goals can lead to cooperative social signals [29, 30]. More proximal intentions are perhaps easier to incorporate within a framework by acknowledging that signallers have goals even if it is as straightforward as a desire to sustain an interaction.

Behind each term in this section is a huge set of descriptive issues. For instance, affective states adequately is a research field in itself [30]. The options at this level are harder to articulate simply than the options for signals, and it is clear that Social Signal Processing needs systematic work on the problem.

3.3. Relationships among variables
3.3.1 Isolated & Combined Intrapersonal Effects.
A common paradigm in older research is to focus on a few selected variables and provide an explanation on how they convey information, how they are interpreted, and ultimately how they influence social behaviour. The resulting models focus on the isolated contributions of a variable. However, those isolated contributions are only simplified pieces of the overarching model. To compare the framework to the English language, each variable is like a letter. On its own, the contribution of
each letter (or variable) to the English language is minimal. The power of the system derives from the way letters can be combined to form words and sentences. Similarly, each variable (or category of variable) is important, but not to the exclusion of the others. To properly model how humans perceive and produce social behaviour, each of the variables must be accounted for, individually and interactively.

During communicative episodes, individuals do not isolate only one variable or category. Instead, information is gathered from all available social signals and is processed, analysed, and interpreted. For example, when having a conversation, nobody focuses only on the other person’s facial expression because that information could be misleading. If the other person is smiling, it could indicate that they are happy. But if that smile is combined with blushing, stuttering, and gazing at their feet, the person is most likely not happy at all, but is trying to mask embarrassment. To properly understand the significance of the variables, a person cannot simply rely on another person’s facial expression, but must also attend to all the relevant signals including the verbal information, the body posture, physiological reactions and so forth. All of these variables are individual and collective signals that humans attend to and process.

A comprehensive computation model would need to address the complexity of the relationship between the variables. Additionally, the variables do not all interact on the same level. Some variables function as a signal and others primarily function as behaviours. Signals and behaviours interact to influence the conceptual level which is the perception and interpretation of the signals and behaviours.

These integrative qualities of the variables have important practical ramifications. In a worst case scenario a signal may be embedded in a minimal increase in intensity of lots of variables. This means each of these variables would need to be assessed if a signal is to be detected. The reality for research projects with limited resources is a selection of a set of variables has to be made, usually guided by practical or historical and discipline related contingencies. As a consequence there may be certain signals that may not be detectable without a broad and inclusive set of variables.

3.3.2 Bidirectional Interpersonal Effects.

To complicate matters further, the bidirectional influence between the individuals engaged in the social interaction must be addressed by the model. While one person is detecting and interpreting the social signals conveyed by the other person, their own social signals are also being detected and interpreted. Social signals emitted are constantly being modified based on these interpretations. The social signals of Person A are influenced by the interpretations of the social signals of Person B, and vice versa [32]. Consequently, a comprehensive model must take into account all the potential social signals of Person A, their interactive relationships, the social signals of Person B, their interactive relationships, and the interactive relationships between Person A’s and Person B’s social signals. With the addition of further people to an interaction, greater sets of social signals and interactive relationships have to be accounted for by a model.

The interdependence of the Person A and Person B’s signals and behaviours not only influence the conceptual mapping, but also the statistical analysis of the framework. The analysis of these interacting variables can be challenging with traditional experimental statistical techniques, and therefore is not appropriate for the analysis of dyadic communication. Instead, other types of statistics (e.g. multilevel modeling) must be applied [32].

Furthermore, a proper model needs to incorporate the ability to attend, perceive, and interpret multimodal signals. Communication is not solely auditory, nor is it solely visual; there are tactile and even olfactory elements. The signals available during communication cross over and interact between modalities. Consequently, the model must account for multimodal perception and attention. Fortunately, multimodal perception and attention is a growing area of research.

3.3.3 Logical relationships

It is tempting to assume that the logical nature of the relationships among variables is a simple conditional – if sign S, then condition C. However, much more complex types of relationship are common, if not the norm. Relationships often abductive (i.e. the best explanation for sign S is that condition C is present) or cancellable (i.e this is my inference about sign S, but new evidence may shows that it is wrong).

Sometimes the logical relationship can be reduced to a simple conditional by considering a complex of signals rather than one in isolation (not only a smile, but one with a particular time course, accompanied by movements of the head and shoulders). It is a key question how far that strategy can be taken.

3.3.4 Computing relationships

Running through the discussion of relationships is a familiar computational issue. It is standard to contrast two approaches to constructing an internal representation of a relationship, conceptually-driven and statistically driven. Conceptually-driven models are based on theoretically proposed structures and relationships. These models have traditionally been fragile. Statistically-driven models involve models derived via machine learning by developing statistical relationships between the structures. These tend to be more robust, but at the expense of conceptual significance, and hence of generalisability. It is not obvious how Social Signal Processing should regard the two options. Some of the disciplines on which it draws are deeply sceptical of statistically-driven analyses, for non-trivial reasons. However, there are obvious practical reasons to use them in many applications.
3.4. Context of communication

3.4.1 Medium.
A framework that includes and models the relationships between all the possible variables needs to be flexible. Some modalities and corresponding variables may not be available during every communicative episode. The medium is a strong determinant of which modalities are present. In a face-to-face conversation, all modalities are usually available. The technologically-mediated means of communication occurs primarily in one modality at the expense of the others. For example, during telephone conversations, verbal exchange and voice tone are present. However, some modalities (e.g. visual system) are no longer providing social signals.

3.4.2 Setting.
The setting refers to the physical characteristics of the surrounding area. This includes the lighting, the space, and the scenery. Being in a darkly lit room results in a degradation of the social signals that are detectable, and this reduction in signal quality can consequently influence behaviour and the interpretation of the communicative episode.

3.4.3 Situation.
This relates to the purpose of the communicative episode instead of its setting. For example, social signals and behaviour will be different during a job interview compared to being in a cinema.

3.4.4 Person by Context Interactions.
Individual and cultural differences affect communicative episodes in two important ways. Firstly, an individual’s ability to accurately detect and interpret social signals is influenced by their personality and their culture. For example, research has shown that individual differences in personality can influence children’s accuracy in the categorisation of emotions [33]. Furthermore, individual and cultural differences will influence the social signals and behaviour emitted by a person. For example, an extremely shy person will produce signals of discomfort during social interactions, whereas his non-shy peer is less likely to do so [34].

The effects of individual differences are also highly dependent on the context. As proposed by Bem & Allen [35], some individuals are more strongly influenced by context, whereas others remain fairly consistent. Recent research in computer-mediated communication (CMC) has highlighted the person by context interaction by demonstrating that some individuals remain consistent in their behaviour across conditions with reduced social signals [36], whereas other can actually benefit from the reduction.

Another major factor is the familiarity and the nature of the relationship between two people. Fewer and subtler social signals can be detected by two people who have a close relationship [37]. For example, siblings can transmit more information with a look or one word than strangers can with an entire conversation.

3.5. Whose interpretation?
A final consideration when developing a computational analysis model is whose interpretation of the signals and behaviour the model is representing. The first consideration is whether the model should reflect how an observer would interpret the signals or should reflect the meaning that the person giving the signals and doing the behaviour intended. Linked to that is the concern that models should be able to adapt to the individual and cultural differences that influence the interpretation. For example, for a British person, a raised index and middle finger with the palm facing in is a vulgar hand gesture. The same gesture for a North American simply represents the number 2.

4. Conclusion
This paper has tried to indicate the range of variables, modalities, relationships, levels, interpersonal differences, contextual effects and interactions that a comprehensive framework needs to accommodate. The next important step is to develop a model that can inform computational analysis the human process of detecting, interpreting and producing social signals and behaviour. The most sensible way to do this is by applying the model to a specific social phenomenon like politeness. By doing so, a model that is both conceptually and statistically driven can be adapted to explain the communicative process of politeness, and then generalized to other communicative processes.

Acknowledgement Preparation of this paper was supported by FP7 projects SEMAINE and SSPnet

References