

# Chapter 2

## Modeling Social Signals and Contexts in Robotic Socially Believable Behaving Systems

Anna Esposito and Lakhmi C. Jain

**Abstract** There is a need for a holistic perspective when considering aspects of natural interactions with robotic socially believable behaving systems, that must account of the cultural, social, physical, and individual (the context) features that shape interactional exchanges. Context (the physical, social and organizational context) rules individual's social conducts and provide means to render the world sensible and interpretable in the course of everyday activities. Contextual aspects of interactional exchanges make any of it unique and requiring different interpretations and actions. A robotic socially believable system must be able to discriminate among the infinities of contextual instances and assign to each their unique meaning. This book reports on the last research efforts in making “natural” human interactional exchanges with social robotic autonomous systems devoted to improve the quality of life of their end-users while assisting them on several needs, ranging from educational settings, health care assistance, communicative disorders, and any disorder impairing either their physical, cognitive, or social functional activities.

**Keywords** Social robotic autonomous systems · (social, physical and organizational) context · End-users' quality of life · Interactional exchanges

### 2.1 Introduction

The human ability to merge information from different sensory systems offers more accurate and faster competencies to operate in response to the environmental stimuli [1]. The integration of different signals in an unique percept is especially

---

A. Esposito (✉)  
Dipartimento di Psicologia and IIASS, Seconda Università di Napoli,  
Caserta, Italy  
e-mail: iiass.annaesp@tin.it

L.C. Jain  
Faculty of Education, Science, Technology & Mathematics,  
University of Canberra, Canberra, Australia  
e-mail: jainlc2002@yahoo.co.uk

© Springer International Publishing Switzerland 2016

A. Esposito and L.C. Jain (eds.), *Toward Robotic Socially Believable Behaving Systems - Volume II*, Intelligent Systems Reference Library 106, DOI 10.1007/978-3-319-31053-4\_2

appreciated in noisy environments with corrupted and degraded signals [20]. Research in neuroscience had proved that audiovisual, visual-tactile and audio somatic sensory inputs are constantly synchronized and combined into a reasoned percept [3, 33]. For example, in speech, the influence of visual on auditory signals perception is proved by the McGurk effect [28]. Recent investigations on the emotional labelling of only audio, mute video, and combined audio/video stimuli proved that the human processing of emotional information is strongly affected by the context to the extent that, depending on the participant' language, culture, and her knowledge of a given foreign language, her performance on an emotion recognition task, accuracy may become significantly worse when the visual signal (emotional facial expressions) is added to audio (emotional vocal expressions, [9, 11–13, 16].

Facial expressions, head, body and arms movements (grouped under the name of gestures) all potentially provide information to the communicative act, supporting the interactional exchange and allowing interactants to add a rich variety of contextual information to their messages including (but not limited to) their psychological state, attitude. Gestures have been shown to vary in size, redundancy, and complexity, depending on the grounding status of the information encoded and/or the meaning attributed to the message. Gestures act in partnership with speech, building up shared knowledge and meanings when the interactional exchange is successful [10, 15, 23].

Psycholinguistic studies have confirmed the partnership nature of verbal and non-verbal signals in human interaction demonstrating that the understanding of a message results from the integration of multi-sensory features appropriately distributed along the interaction [30].

In addition, it has been suggested that interactive communication is emotionally driven [8] and that the encoding and decoding procedures exploited by humans to express and/or read emotions are fundamental to secure the social quality and the cognitive functioning of a successful interactional exchange.

Another crucial aspect of multimodal communication is the relationship between paralinguistic and extra-linguistic information (such as speech pauses, head nodding). Psycholinguistic studies have shown that there exists a set of non-lexical expressions carrying specific communicative values (expressing for example turn-taking and feedback mechanism regulations) such as empty and filled pauses and other hesitation phenomena [14, 15, 17]. It has been shown that pauses (holds) in gestures plays similar communicative values and synchronize with paralinguistic information [10].

It can be concluded that the verbal and nonverbal communication modes jointly cooperate in assigning semantic and pragmatic contents to the conveyed message by unraveling the participants' cognitive and emotional states and exploiting this information to tailor the interactional process. These modes exploit multimodal social signals and are tailored to the contextual instance in which they are explicated.

The huge demand for complex autonomous systems able to assist people on several needs had produced a consistent number of EU and overseas funded projects such as (a) ERICA ([www.jst.go.jp/erato/ishiguro/en/](http://www.jst.go.jp/erato/ishiguro/en/)) a conversational android with human appearance aiming to interact with humans through multimodal social signals such as face, speech, and body movements; (b) MUMMER a humanoid robot

(based on Aldebaran's Pepper platform, [http://www.dcs.gla.ac.uk/vincia/?page\\_id=116](http://www.dcs.gla.ac.uk/vincia/?page_id=116)) engaging people in dynamic environments, such as shopping malls; (c) TESLA—An Adaptive Trust-based e-assessment System for Learning ([www.open.ac.uk/iet/main/research-innovation/research-projects/adaptive-trust-based-e-assessment-system-learning-tesla](http://www.open.ac.uk/iet/main/research-innovation/research-projects/adaptive-trust-based-e-assessment-system-learning-tesla)); (d) ALIZ-E (<http://www.aliz-e.org/>) engaging diabetic children in a series of real-world situations (partially) WOZ-simulated implemented using the NAO robot (<http://www.aldebaran-robotics.com/en>) [25, 26]; and several more, all exemplifying the huge research efforts in implementing socially believable assistive technologies. However, these projects have been characterized by a scarce care to what effectively would have been end-users' requirements and expectations to qualify as "socially behaving" agents providing "*social/physical/psychological/assistive ICT services. In particular, the term "social robotics" envisions a "natural" interaction of such devices with humans, where "natural" is interpreted as the ability of such agents to enter the social and communicative space ordinarily occupied by living creatures*" [7, p. 2]. In addition really few has been made to account for "*how the interaction between the sensory-motor systems and the inhabited environment (that includes people as well as objects) dynamically affects/enhances human reactions/actions, social perception and meaning-making practices*" [6, p. 6].

To account for the above mentioned problems it is needed an all-embracing prospect pushing the designer to contemplate the system's behavior and appearance (in order for the system to be social), the trustworthiness the user put into it (in order to be emphatic) taking into account the contextual instance (the scenario) and the system' functionalities required in each situation, as well as, the individual's social rules and cognitive competencies [5].

In the field of Human Robot Interaction, such an approach, will require investigations on the cognitive architectures and cognitive integrations needed for accounting of human behavior across different domains, and inherently of the behavior humans engage with a system that, as much as complex and autonomous can be, can offer only a sub-optimal interaction process (see key activities of the topic group Natural Interaction with Social Robots <http://homepages.stca.herts.ac.uk/~comqkd/TG-NaturalInteractionWithSocialRobots.html>, [6, 7, 19]).

To date there have been relatively few efforts assessing human interactional exchanges in context in order to develop complex autonomous systems able to detect user's trust and mood, rise emphatic feeling, and take actions to provide help. In addition there are no standards for the development of more 'satisfying' complex autonomous systems that account for user's expectations and requirements in a structured manner. Although there have been efforts in providing suggestions for potential solutions [2, 18, 32] this issue is at a research stage [7, 19].

Generally the development and assessment of complex autonomous systems is tackled using two different approaches: user's self reports and performance based measures. The user's self reports can be highly criticized because of the user's difficulty to accurately describe her/his expectations and requirements, being them generally technologically naïve and suspicious. Performance-based measures can be considered more reliable since they require the execution of specific tasks assessed

by a trained evaluator. Nevertheless, these tasks are generally carried out under artificial conditions and require extensive equipment, well defined environmental context and time consuming evaluation procedures that do not value the daily spontaneous activity producing biases in the collected measures. Despite its importance, generally these systems are unable to being context-aware, to adapt to user's preferences and very distinct needs and to correctly interpret all user's actions.

The research papers proposed in this book investigate the features that are at the core of human interactions and provide attempts to model the cognitive and emotional processes involved in order to design and develop complex autonomous system prototypes able to simulate the human's ability to decode and encode social cues while interacting.

## 2.2 Content of the Book

The research objectives proposed in this book can be interpreted as a meta-methodology aimed to investigate features that are at the core of human interactional exchanges and model the cognitive and emotional processes involved in interactions, in order to design and develop complex autonomous system prototypes able to socially behave in (at the least) specific scenarios. The attention is focused on the analysis and modeling of social behavioral features and human ability to decode and encode social cues while interacting. Behavioral data (speaking, body movements, facial, vocal and gestural emotional expressions) are gathered from healthy and communicative or socially impaired participants. This require the definition of behavioral tasks that serve both to detect changes in the healthy, as well as, impaired perception of social cues. Specific scenarios are proposed for these tasks intended to assess the users' attitude, acceptance, and trustworthiness toward a robotic system considering its emphatic and social competencies, as well as appearance. The collected data are used to gain knowledge on how behavioral and interactional features are affected by individual characteristics and personalities, contextual instances, and environmental perceptual features. Hopefully, these investigations will guide on which human-like social characteristics and appearance (physically embodied or virtual intelligent agent?) a complex autonomous system should exhibits to gain the users' trust and acceptance as a socially behaving agent.

To this aim, the book includes nine investigations on the mathematical modeling of social signals and context embedded in interactional exchanges. The second chapter by Maldonato and Dell'Orco [27] affords one of the most debated issue in artificial intelligence: "*the possibility of reproducing in an artificial agent (based on formal algorithms) some typically human capacities (based on natural logic algorithms) such as consciousness, the ability to deliberate and make moral judgments*" [p. 1]. The authors are very desecrators to the point of asserting that "*clarifying consciousness mechanisms of artificial organisms could help us to discover what we still ignore about neurobiological consciousness*" [p. 2]. This extreme in the quest for equipping machines with human level automaton intelligence, consciousness,

and intuition, leave us with the question on whether we really want conscious and intuitive artificial agents. The answer is given by the contribution of Gnjatović and Borovac [21] which propose the implementation of conscious-like conversational agents, implicitly answering to the question with a discussion on which are the limitations for implementing consciousness features in mathematical prototypes. These investigations clearly suggest that the aims of the research on social robotics is to implement natural interactions with such social agents. The contribution of Vogel [35] covers one aspect of this sociability proposing an investigation aimed to the “*understanding of natural dialogue*” in order to “*fully inform the construction of believable artificial systems that are intended to engage in dialogue with a manner close to human interaction in dialogue*” [p. 1]. Harrington et al. [22] follow a similar vision considering “*the relevance of context and experience for the operation of historical sound changes*” [p. 1]. The contribution of Clavel et al. [4] is an original survey on which competencies an artificial agent must exploit to maintain, when interacting, users’ engagement. The focus is on both users’ attentional and emotional involvement. op den Akker’s et al. [31] contribution propose “*Kristina, a personal digital coaching system built to support and motivate users to live a balanced and healthy lifestyle*” [p. 14]. The authors are aware of dangers and objections that can be raised by modeling interactional persuasive features and provide a very interesting discussion on these aspects. The contribution of Vinciarelli [34] discuss on how “*endowing machines with social perception*”, in particular “*by providing a simple conceptual model of social perception and by showing a few examples related to Automatic Personality Perception, the task of predicting how people perceive the personality of others*” [p. 1]. Finally, the last two contributions surpass dyadic interactional features considering to model either multimodal and multiparty interactions in educational settings, as in the work of Koutsombogera et al. [24], or to detect abnormal behavioral patterns in crowd scenarios, as in the contribution of Mousavi et al. [29].

## 2.3 Conclusions

The readers of this book will get a taste of the major research areas in modeling social signals and contextual instances of interactional exchanges in different scenarios for implementing robotic socially believable behaving systems. This research should result in a series of theoretical and practical advances in the field of cognitive, and social psychology such as: (1) Repertories of social signals better illustrating the cognitive, semantic, emotional and semiotic mechanisms essential for successful interactional human-machine exchanges; (2) Models for representing data, reasoning, learning, planning, and decision making, as well as, individual/group behavior analysis models in multilingual and cross-cultural contexts; (3) The identification of new interactional persuasive and affective strategies and contextual instances calling for their use. Considering technological issues, the present research must lead to (1) New computational approaches and departures from existing cognitive frameworks and

existing algorithmic solutions such as dynamic Bayesian networks, long short-term memory networks, and fuzzy models of computation; (2) The implementation of behaving ICT systems of public utility and profitable for a living technology that simplifies user access to future, remote and nearby social services encompassing language barriers and cultural specificity; (3) Market applications such as: context-aware avatars replacing human in high risk tasks, companion agents for elderly and impaired people, socially believable robots interacting with humans in extreme, stressful time-critical conditions, future smart environments, ambient assistive living technologies, computational intelligence in games/storytelling, embodied conversational avatars, and automatic healthcare and education services.

## References

1. Block N (1995) The mind as the software of the brain. In: Smith EE, Osherson DN (eds) *Thinking*. MIT Press, Cambridge, pp 377–425
2. Brandão Moniz A (2010) Anthropocentric-based robotic and autonomous systems: assessment for new organisational options. IET Working Papers Series No. WPS07/2010
3. Callan DE et al (2003) Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* 14:2213–2218
4. Clavel C, Cafaro A, Campano S, Pelachaud C (2016) Fostering user engagement in face-to-face human-agent interactions: a survey. This volume
5. Davis MH (1983) Measuring individual differences in empathy: evidence for a multidisciplinary approach. *J Personal Soc Psychol* 44:113–126
6. Esposito A, Esposito AM, Vogel C (2015) Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognit Lett* 66:41–51
7. Esposito A, Fortunati L, Lugano G (2014) Modeling emotion, behaviour and context in socially believable robots and ICT interfaces. *Cognit Comput* 6(4):623–627
8. Esposito A (2013) The situated multimodal facets of human communication. In: Rojc M, Campbell N (eds) *Coverbal synchrony in human-machine interaction*, ch. 7. CRC Press, Taylor & Francis Group, Boca Raton, pp 173–202
9. Esposito A, Esposito AM (2012) On the recognition of emotional vocal expressions: motivations for an holistic approach. *Cognit Process* 13(2):541–550
10. Esposito A, Esposito AM (2011) On speech and gesture synchrony. In: Esposito A et al (eds) *Communication and enactment*. LNCS, vol 6800. Springer, New York, pp 252–272
11. Esposito A, Riviello MT (2011) The cross-modal and cross-cultural processing of affective information. In: Apolloni B et al (eds) *Frontiers in artificial intelligence and applications*, vol 226. IOSpress, Amsterdam, pp 301–301
12. Esposito A (2009) The perceptual and cognitive role of visual and auditory channels in conveying emotional information. *Cognit Comput J* 1(2):268–278
13. Esposito A, Riviello MT, Bourbakis N (2009) Cultural specific effects on the recognition of basic emotions: a study on Italian Subjects. In: Holzinger A, Miesenberger K (eds) *USAB 2009*. LNCS, vol. 5889. Springer, Berlin, pp 135–148
14. Esposito A (2008) Affect in multimodal information. In: Tao J, Tan T (eds) *Affective information processing*. Springer, Heidelberg, pp 211–234
15. Esposito A, Marinaro M (2007) What pauses can tell us about speech and gesture partnership. In: Esposito A et al (eds) *Fundamentals of verbal and nonverbal communication and the biometric issue*. NATO series human and societal dynamics, vol 18. IOS press, The Netherlands, pp 45–57

16. Esposito A (2007) The amount of information on emotional states conveyed by the verbal and nonverbal channels: some perceptual data. In: Stilianou Y et al (eds) *Progress in nonlinear speech processing*. LNCS, vol 4391. Springer, Heidelberg, pp 245–268
17. Esposito A (2006) Children's organization of discourse structure through pausing means. In: Faundez\_Zanuy M et al (eds) *Nonlinear analyses and algorithms for speech processing*. LNCS, vol 3817. Springer, New York, pp 108–115
18. Feil-Seifer D, Skinner K, Mataric MJ (2007) Benchmarks for evaluating socially assistive robotics. *Interact Stud: Psychol Benchmarks Hum-Robot Interact* 8(3):423–429
19. Fortunati L, Esposito A, Lugano G (2015) Beyond industrial robotics: social robots entering public and domestic spheres. *Inf Soc: Int J* 31(3):229–236
20. Garrigan P, Kellman PJ (2008) Perceptual learning depends on perceptual constancy. *PNAS* 105(6):2248–2253
21. Gnjatović M, Borovac B (2016) Toward conscious-like conversational agents. This volume
22. Harrington J, Kleber F, Reubold U, Stevens M (2016) The relevance of context and experience for the operation of historical sound change. This volume
23. Kendon AG (2005) *Visible action as utterance*. University Cambridge Press, New York
24. Koutsombogera M, Deligiannis M, Giagkou M, Papageorgiou H (2016) Towards modelling multimodal and multiparty interaction in educational settings. This volume
25. Kruijff-Korbayová I, Athanasopoulos G, Beck A, Cosi P, Cuayahuitl H, Dekens T, Enescu V, Hiolle A, Kiefer B, Sahli H, Schröder M, Somlavilla G, Tesser F, Verhelst W (2011) An event-based conversational system for theNAO robot. In: *IWSDS 2011*. Granada, Spain
26. Kruijff-Korbayova I, Cuayahuitl H, Kiefer B, Schröder M, Cosi P, Paci G, Somlavilla G, Tesser F, Sahli H, Athanasopoulos G, Wang W, Enescu V, Verhelst W (2012) Spoken language processing in a conversational system for child-robot interaction. In: *Workshop on child-computer interaction*
27. Maldonato M, Dell'Orco S (2016) Adaptive and evolutive algorithms: a natural logic for artificial mind. This volume
28. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748
29. Mousavi H, Galoogahi HK, Perina A, Murino V (2016) Detecting abnormal behavioral patterns in crowd scenarios. This volume
30. Munhall KG, Jones JA, Callan DE, Kuratate T, Vatikiotis-Bateson E (2004) Visual prosody and speech intelligibility. *Psychol Sci* 15(2):133–137
31. op den Akker HJA, Klaassen R, Nijholt A (2016) Virtual coaches for healthy lifestyle. This volume
32. Petrelli D, Not E (2005) User-centred design of flexible hypermedia for a mobile guide: reflections on the hyperaudio experience. *User Model User-Adapt Inter* 15(3–4):303
33. Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cognit Neurosci* 19(12):1964–1973
34. Vinciarelli A (2016) Social perception in machines: The case of personality and the Big-Five traits. This volume
35. Vogt C (2016) Communicative sequences and survival analysis. This volume



<http://www.springer.com/978-3-319-31052-7>

Toward Robotic Socially Believable Behaving Systems -

Volume II

Modeling Social Signals

Esposito, A.; Jain, L.C. (Eds.)

2016, XIV, 207 p. 45 illus., 17 illus. in color., Hardcover

ISBN: 978-3-319-31052-7