Engagement in Human-Agent Interaction: An Overview

Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir,
Mohammad Obaid, Catherine I Pelachaud, Christopher Peters

Engagement in Human-Agent Interaction: An Overview

Catharine Oertel [1]*, **Ginevra Castellano** [2], **Mohamed Chetouani** [3], **Jauwairia Nasir** [4], **Mohammad Obaid** [5], **Catherine Pelachaud** [6], **and Christopher Peters** [7]

[1] *Interactive Intelligence, Intelligent Systems, EWI, Delft University of Technology, Delft, Netherlands*
[2] *Uppsala Social Robotics Lab, Department of Information Technology, Uppsala University, Uppsala, Sweden*
[3] *Institute for Intelligent Systems and Robotics, CNRS UMR 7222, Sorbonne University, Paris, France*
[4] *CHILI Lab, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[5] *UNSW Art & Design, University of New South Wales, Sydney, NSW, Australia*
[6] *CNRS, Institute for Intelligent Systems and Robotics, Sorbonne University, Paris, France*
[7] *Embodied Social Agents Lab (ESAL), School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden*

Correspondence*:
Catharine Oertel
C.R.M.M.Oertel@tudelft.nl

2 **ABSTRACT**

3   Engagement is a concept of the utmost importance in human-computer interaction, not only for
4 informing the design and implementation of interfaces, but also for enabling more sophisticated
5 interfaces capable of adapting to users. While the notion of engagement is actively being studied
6 in a diverse set of domains, the term has been used to refer to a number of related, but different
7 concepts. In fact it has been referred to across different disciplines under different names and
8 with different connotations in mind. Therefore, it can be quite difficult to understand what the
9 meaning of engagement is and how one study relates to another one accordingly. Engagement
10 has been studied not only in human-human, but also in human-agent interactions i.e. interactions
11 with physical robots and embodied virtual agents. In this overview article we focus on different
12 factors involved in engagement studies, distinguishing especially between those studies that
13 address task and social engagement, involve children and adults, are conducted in a lab or aimed
14 for long term interaction. We also present models for detecting engagement and for generating
15 multimodal behaviors to show engagement.

Table 1. This table illustrates a selected list of articles, based on Section 3, to represent our annotation procedure and to provides an overview of the attributes annotated including embodiment, role, affect, participants, study design, type of interaction capabilities, and measurements.

Article	Embodiment		Role									Affect	Participants				Study						Type			Measurements							
	Robot	Virtual Agent	Host	Game Player	Instructor/Tutor	Storyteller	Companion	Entertainer	Conversational Partner	Interviewer	Multiple Roles	Included	Children	Adults	< 30	≥ 30	Lab Study	Real-time	Autonomous	WoZ	Sensors	Perception	Generation	Perception	Simulation	Questionnaires	Video	Dialogue	Contextual	Tracking Tech.	Observations	Physiological	Time
Sidner and Dzikovska (2002)	●	○	●	○	○	○	○	○	○	○	○	○	●	●	–	–	◐	●	●	○	●	●	●	●	○	○	●	○	○	○	●	○	○
Alami et al. (2005)	●	○	○	○	●	○	○	○	○	○	○	○	○	●	–	–	–	●	●	○	●	●	●	●	●	○	●	●	●	○	●	●	○
Nakano and Nishida (2005)	○	●	○	○	○	●	○	○	○	○	○	○	○	●	–	–	●	●	●	○	●	●	●	○	○	○	○	○	●	○	●	○	○
Castellano et al. (2009)	●	○	○	●	○	○	○	○	○	○	○	●	●	○	●	○	○	●	●	○	●	●	○	●	○	○	●	○	○	●	○	○	○
Bohus and Horvitz (2009b)	○	●	●	○	○	○	○	○	○	○	○	○	○	●	●	○	○	●	●	○	●	●	○	●	○	○	●	●	●	●	○	●	○
Szafir and Mutlu (2012)	●	○	○	○	○	●	○	○	○	○	○	–	○	●	○	●	●	–	–	–	●	●	●	●	○	●	○	○	○	○	○	●	○
Vázquez et al. (2014)	●	○	○	○	○	○	○	●	○	○	○	●	●	○	○	●	◐	●	○	●	●	●	○	●	○	○	●	●	○	○	●	○	○
Moshkina et al. (2014)	●	○	○	●	○	○	○	○	○	○	○	○	●	○	○	●	○	●	●	○	●	●	●	○	○	○	●	○	○	○	○	●	○
Rehm and Jensen (2015)	○	●	○	○	●	○	○	○	○	○	○	●	●	○	○	●	○	●	●	○	●	●	●	●	○	●	●	○	○	○	○	●	○
Anzalone et al. (2015)	●	○	○	○	○	○	○	○	○	●	○	●	●	●	–	–	●	●	●	○	●	●	●	●	○	○	○	○	○	●	○	○	○
Subramainan et al. (2016a)	○	●	○	○	●	○	○	○	○	○	○	●	●	○	–	–	●	○	○	○	–	–	○	○	●	–	–	–	–	–	–	–	–
Salam et al. (2017)	●	○	○	○	○	○	○	●	●	○	○	○	○	●	●	○	○	●	○	●	●	●	○	●	○	○	●	●	○	○	○	○	○
Trinh et al. (2018)	○	●	○	○	○	○	○	○	●	●	○	○	○	●	●	○	○	●	●	○	●	●	○	●	○	○	●	●	○	○	○	○	●
Inoue et al. (2018)	○	●	○	○	○	○	○	○	○	●	○	●	○	●	●	○	●	●	○	○	●	●	○	●	○	○	●	●	●	○	○	○	○
Sohail et al. (2019)	○	●	○	○	○	●	○	○	○	○	○	●	○	●	●	○	●	●	●	○	●	●	○	●	○	●	●	●	○	○	○	○	○
Biancardi et al. (2019a)	○	●	○	○	○	○	○	●	○	○	○	●	○	●	●	○	●	●	●	○	●	●	●	●	○	●	●	●	○	○	○	○	○
Youssef et al. (2019)	●	○	○	○	○	○	○	●	○	○	○	●	○	●	○	●	○	●	●	○	●	●	○	●	○	●	●	●	○	●	○	○	●

●=Yes; ○=No; " – "=N/A; ◐=In-between.

1   INTRODUCTION

Engagement is a concept of the utmost importance in human-machine interaction, not only for informing the design and implementation of interfaces, but also for enabling more sophisticated interfaces capable of adapting to users. This is particularly true when the interface is an agent (see Figure 1), be it virtual or robotic, that converses with human users. In the former case, the agents detect users' engagement (see Figure 2) while in the latter case the agents adapt to the detected engagement. These agents all have a common goal, namely to have users continue interacting with them and thus manage users' engagement in the interaction. Thus, for human-agent interaction engagement, both perception and generation are important issues. Perceiving how engaged users are can be beneficial information for adapting agent behaviour. It can also be a sign of the quality of the interaction and user's experience with the system. Similarly generating engaged behaviours in an agent can be beneficial for human-perception in terms of social awareness.

The term engagement is being used across a number of diverse research domains, both scientific and commercial. Its definition and use varies considerably and can be confusing, especially for researchers approaching the topic for the first time. In fact engagement recently gained increasing popularity, in particular with the development of interaction paradigms between humans and embodied agents, such as virtual characters and robots (see for example Leite et al. (2016)). Yet, there remains great variability, overlap and often vagueness with respect to the definition of engagement. It is often used synonymously to refer to a number of related concepts, such as interest, sustained attention, immersion and involvement. Recently several papers attempted to provide definition(s) of engagement. Researchers have also proposed computational models to compute engagement, both to analyse a human's level of engagement and to drive agent's behaviour to show its engagement. See Sanghvi et al. (2011) and Sidner et al. (2003) for respective examples. The models vary in terms of definition of engagement (which phenomenon is modeled) and of expressive manifestation (which multimodal behaviours are involved). The fact that more and more papers are trying to provide an overview of engagement is warranted by the great diversity of definitions of engagement across different papers. *The purpose of this article is to provide insights into the use of engagement, particularly as it pertains to human-agent interaction, with a focus on embodied agents such as embodied conversational agents and social robots.*

In this article, we review the literature with the aim of answering the following questions:

(1) Specifically for human-agent interaction, what are the engagement definitions most commonly used and how do they differ from one another?

(2) How does the definition of engagement and its implementation differ along several factors such as interaction settings (real world or laboratory), interaction types (short or long interaction), interaction goals (social interaction or task performance) and user types (adults or children)?

(3) How are engagement annotations being conducted? Which methods and features are being used to detect engagement and which expressions are being used to generate engagement behaviours?

(4) What are the functions of engagement and which adaptation strategies are applied by the agent in order to maintain it, increase it or show disengagement?

The structure of the paper is as follows. First we present our method to gather articles in the literature. We present the annotation schema we follow to analyse and cluster these papers. In Section 3 we list the various definitions of engagement found in the selected papers. From the next section onward, each section focuses on specific aspects of engagement. In Section 4 we review how engagement is defined through the

different scenarios that are commonly used in human-agent interaction. Section 5 focuses on computational models to detect engagement while Section 6 describes the models to drive agent's behaviours to display engagement. Section 7 considers the additional issues and requirements when humans are to engage in sustained interaction over long periods of time with artificial systems. Since many of the studies considered here have adult participant groups, in Section 8 we specifically report those studies that involve children. Finally, we conclude by highlighting gaps in the literature and point to possible future research.

2 METHOD

We started curating a list of relevant papers, in November 2017, by doing a search query using the terms "engagement+human-robot interaction" in Scopus and Google Scholar, two popular citation databases. Thereafter, to consolidate a full list of any newly published articles, the search was repeated every 6 months up until December 2019. The curated list of papers went through the following inclusion criteria:

- Is the paper covering the topic of engagement in human-agent interaction, including detecting it and generating behavior to manifest engagement? This criterion entails that papers can include robotic agents or virtual agents.
- Is the paper over four pages in length? With this criterion, the inclusion of abstracts or poster publications is eliminated.
- Is the content of the paper not overlapping significantly with another paper from the same author(s)? If the paper does overlap, then the most elaborated paper is selected for inclusion.

The initial curated list resulted in 189 papers in total and, based on the inclusion criteria above, 20 were excluded resulting in a final set of 169 papers. Based on a preliminary review of the 169 papers and the questions enumerated in the Introduction section, we developed an annotation schema (see Table 2) to allow us investigate and answer each of the questions.

Using the selected 169 relevant articles, we conducted a full review on each paper to extract the details of the annotation schema categories presented in Table 2. The procedure commenced with an assignment of annotation task to authors of this paper. During the process, the annotators discussed and resolved any ambiguous statements in papers that relate to any of the annotation categories. In order to facilitate the reviewing process and to guarantee reproducibility, a shared spread-sheet was created containing all annotated data. The schema categories represent factual data extracted from each paper. While all 169 papers contribute to the general overviews, discussions and statistics in this survey, due to space constraints, it is not possible to report in detail on all of them. Therefore, all papers are included in the references section, but only a subset of those papers are cited when they have been discussed in more detail.
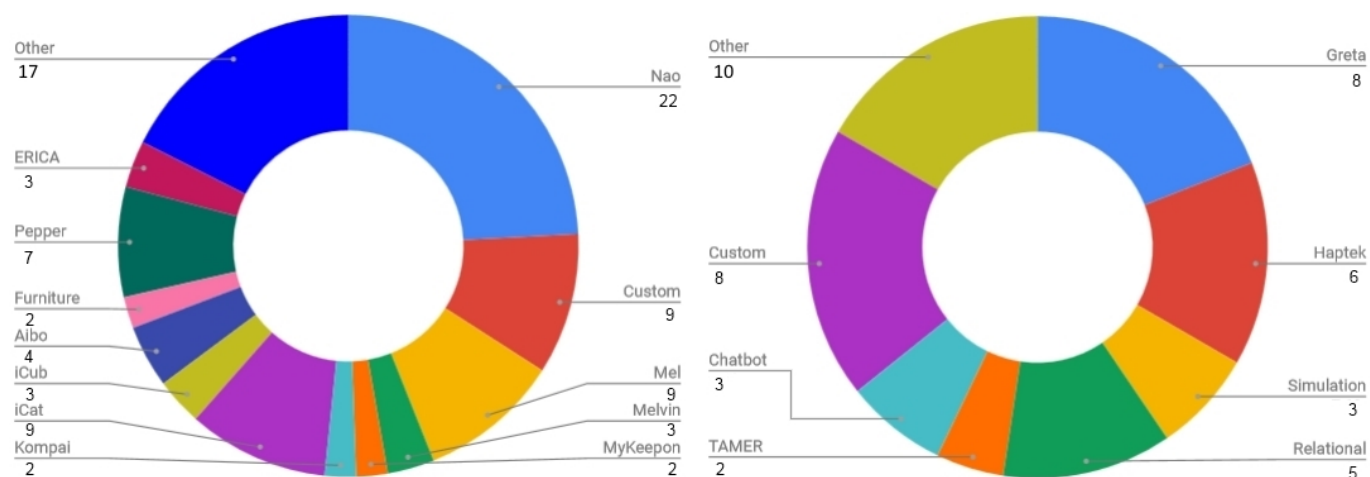

Figure 1. Examples of virtual and physical agents in typical engagement scenarios with humans.
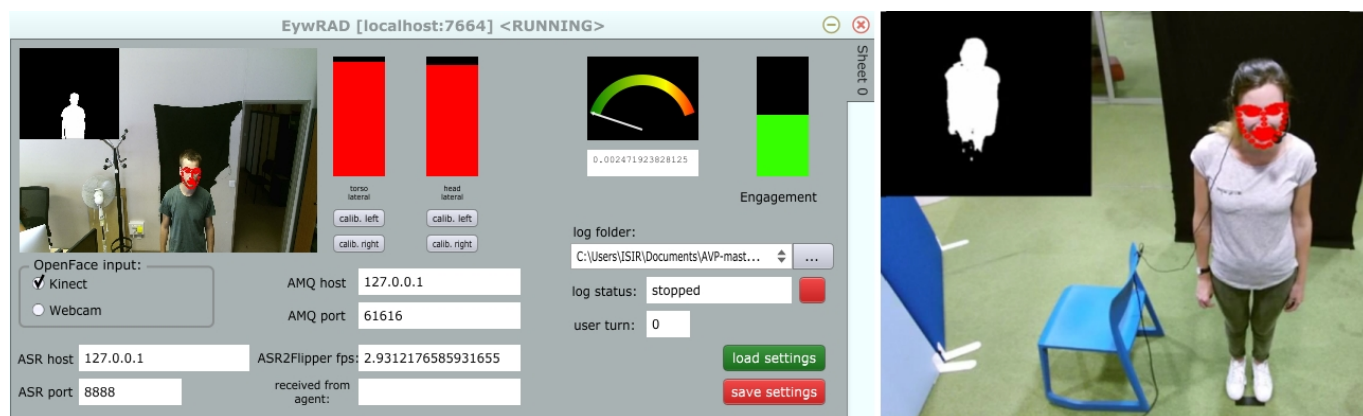
Figure 2. Examples of engagement detection systems.
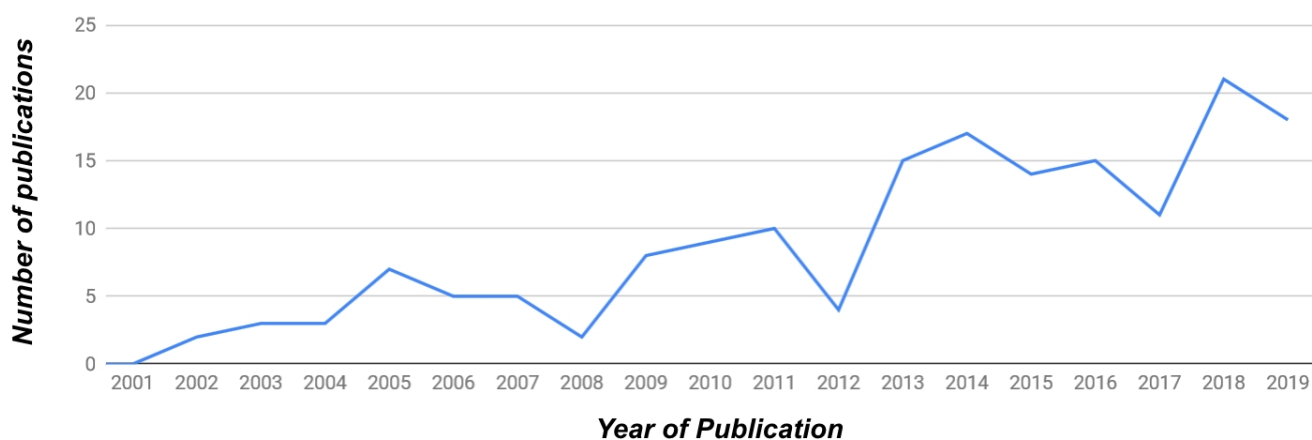
Figure 3. The distribution of publications per year covered in this survey. A total of 169 publications were considered in total between the years 2001 and 2019.

Figure 4. Overview of number of mentions of specific (left) robot types and (right) virtual agent types used in studies. Note that some studies involved the use of multiple robot/agent types, while others did not use any robot or virtual agent, or did not specify the type involved.

| Category	Description
Title	The title of the research paper or article
Year	Publication Year
Scenario	A description of the human-robot scenario used in the article
Application	What is the main application domain presented in the article
Robot	Type of robot(s) presented in the article
Role	The main role of the robot/agent with in the scenario presented
Definition	The definition of 'engagement' presented in the context of the article
Affect	Is there an affective component to the engagement definition
Participants' Type	Are the participants of the study children, adults or both
Number of Participants	Indicates the number of participants included in any study (or studies) presented in the article
Lab Study	Did the article include a study that is conducted in a laboratory settings
Sensors used perception/how?	What sensors are used to measure engagement (if any)
Perception	Is sensory information used for training an ML model?
Research Objectives	What are the aims (purpose), research questions and/or hypotheses presented in the article
Engagement Type	Is the article focusing on the perception, generation or simulation of engagement
Study attributes	What are the main evaluated attributes presented in the article
Scenario Tag	In what context is the interaction taking place?
Finding/s	The main finding presented in the article
Mode	Was the operating mode of the robotic agent automated or based on a Wizard-Of-Oz approach?
Language	What language was the study conducted in?
Country	Where was the study conducted?
Perception measurements	How is participant perception measured? Questionnaires/video analysis/observation
Mode of use	Was the technical setting conducted in real-time, simulated or crowd sourced
Type of paper	What is the main type of the article presented (e.g. Technology, User evaluations etc.)

Table 2. A summary of the annotation schema categories used.

2.1 Statistics

See Figure 3 for a graph of the number of publications covered in this survey according to year. Overall, a total of 169 publications were considered between the years 2001 and 2019. Of the 169 papers, 139 concerned a physical robot (88 papers, i.e. 52% of overall papers) and/or virtual agent (51, 30%) embodiment. 39 (23%) papers involved studies that included or focused on children as participants.

Papers were categorised according to the application area of the paper, the role of the agent/robot, type of robot, the number of participants in the study and the age of participants (adults or children). For each specific category, labels were enumerated by surveying all included papers and then maintaining those labels that appeared in more than one paper. See Figure 4 for a graph of the different robot embodiments

98   used in studies covered in this survey. The labels consisted of a type or name of robot or virtual agent if
99   more than one study took place with it. Otherwise the label *other* was used. *Greta*, *Relational*[1], *Haptek*
100 and *TAMER* refer to specific virtual agent solutions or frameworks, while *Chatbot*, *Simulation*, *Custom*
101 and *Other* are categories. A graph of the applications related to engagement covered in the papers in
102 this survey is presented in Figure 5. They cover the categories of *education, health, games, companion,*
103 *conversation, sales, therapy, host* and *collaboration*. See Figure 6 for a graph of the role of the agent.
104 The role category included the labels *assistant, storyteller, game opponent, demonstrator, collaborator,*
105 *teleoperated, tutor, therapist, elicitor, entertainer, learner, instructor, conversation partner, persuader, host,*
106 *interviewer, contact seeker, motivator, audience, multiple* (for multiple categories) and *other*. A few of the
107 less well-known roles can be better explained by examples; for instance, in Anzalone et al. (2015), the
108 robot is used as an *elicitor*, i.e., it elicits certain behaviors in humans in various face-to-face interaction
109 scenarios. Furthermore, in Rani and Sarkar (2005), a *teleoperated* robot detects the engagement levels of
110 its operator through physiological sensors and adapts its behavior accordingly. Lastly, in Baek et al. (2014),
111 where the robot acts as a *contact seeker*, a study is conducted to explore how a communicator type (human,
112 robot, product) impacts social presence and shyness of participants when they come in physical contact
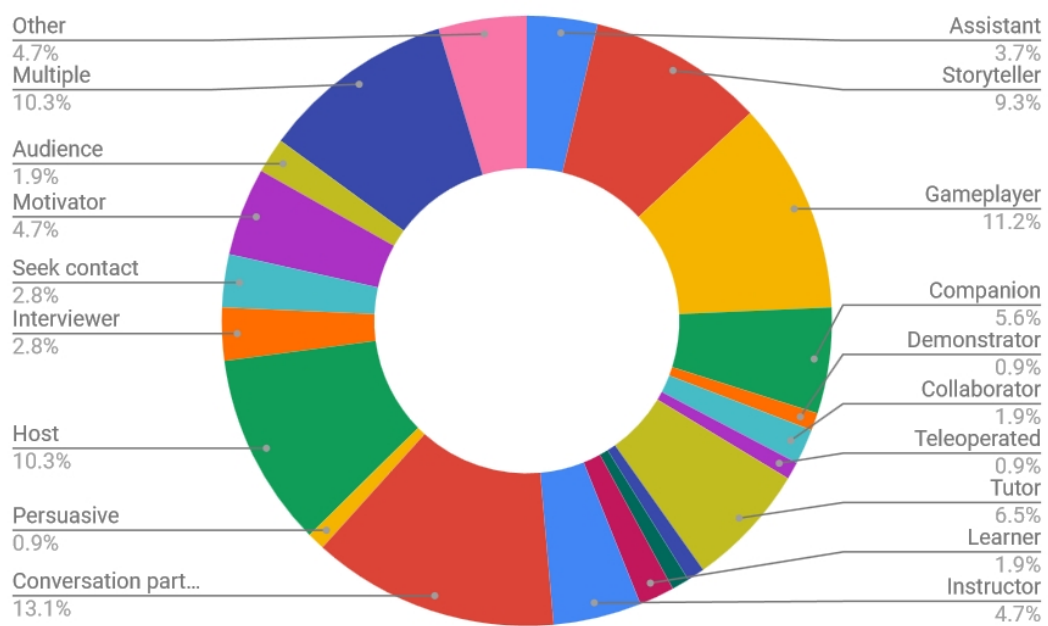113 with each of them.
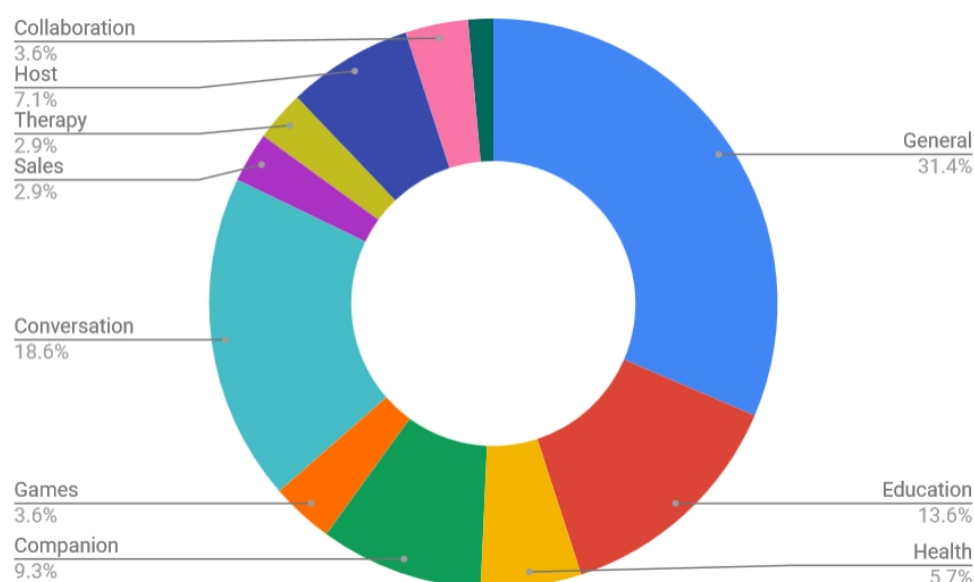
Figure 5. Overview of all publications in this survey according to application type when it was specified (140 in total specified, 29 unspecified or could not be identified).

114

2.2   Evaluation Methodology

116   To understand how evaluations in the curated list of papers investigated the users' perception in their
117 work, we extracted the type of evaluation data collected in each of the papers. After an initial review of
118 all the papers, it was apparent that 23 papers did not include an evaluation study and were found to be
119 focused on technical, modelling or conceptual contributions. For the remaining papers, we found thirteen
120 types of evaluation tools, both objective and subjective, to measure the users' perception in the presented
121 works. The data types reported include questionnaires, RGB video recordings, depth camera recordings,

---

Figure 6. Overview of all publications in this survey according to the role of the robot or virtual agent when it was specified (107 in total specified, 62 unspecified or could not be identified).

122 time/temporal performances, post study interviews, observations, physiological sensor data, tracking
123 sensor technologies (motion, eye and laser tracking), speech and dialogue recording, and contextual and
124 application records (such as game scores, number of moves, implicit touch gestures, logs etc). Figure 7
125 represents the percentage of each type used in the curated list of papers.
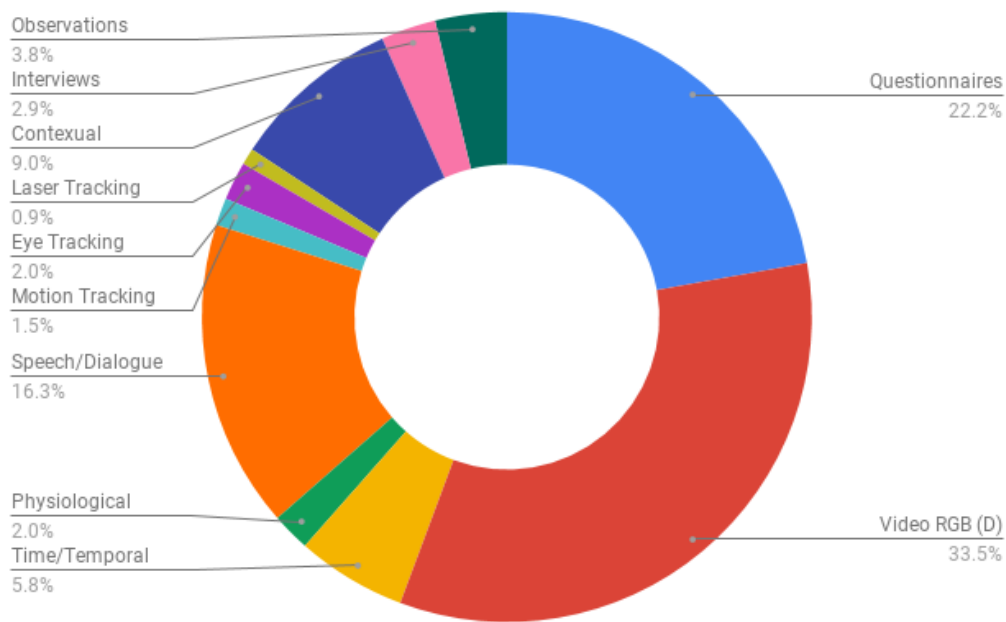
Figure 7. The type of data collected in the evaluation studies conducted in the curated list of papers. Note that many studies collected multiple data types while some did not collect any.

126   It is apparent from Figure 7 that authors approached measuring the users' perception in HRI engagement
127   research mainly through video analysis (33.5%), speech/dialogue analysis (16.3%), contextual and
128   application performance data (9.0%) and questionnaires (22.2%).

3   DEFINITION OF ENGAGEMENT

129   As previously mentioned, there is great variation concerning definitions of engagement. Papers can be
130   divided into those that consider engagement as a process and those that treat engagement as a state. The
131   state point of view assumes that one is either engaged or not engaged (e.g. Inoue et al. (2018)), while the
132   process point of view assumes that there are different processes that unfold during an interaction. Here
133   the action of getting engaged is part of the construct of engagement itself. The most commonly used
134   example of a process definition is in Sidner et al. (2003), which defines engagement as 'the process by
135   which interactors start, maintain, and end their perceived connections to each other during an interaction'.
136   Examples of studies which are using this definition are Holroyd et al. (2011); Bohus and Horvitz (2009b);
137   Alami et al. (2005); Sidner et al. (2006); Nakano and Nishida (2005); Anzalone et al. (2015). There are also
138   those who slightly adapt or alter the definition. One example is Bohus and Horvitz (2009a). Their definition
139   of engagement is "the process subsuming the joint, coordinated activities by which participants initiate,
140   maintain, join, abandon, suspend, resume, or terminate an interaction". It also includes the concepts of
141   abandon, suspend and resume.

142   A second distinction can be made depending on who or what is the receiver of user engagement. For
143   example, in human-agent interaction, the human user can be engaged with the agent (and vice-versa), the
144   task that user and agent are be involved in, or the whole system (i.e., agent and task). The former case is
145   often called *social engagement* and the next one *task engagement*.

146   Regarding the definition of engagement in a conversational setting, Coker and Burgoon (1987) were the
147   first, to our knowledge, who attempted a definition. They are referring to a concept called *conversational*
148   *involvement* which, for all intents and purposes of this paper, refers to the same concept as engagement.
149   They defined four distinct variables: "the degree of animation and dynamism", "the tendency to be interested
150   in, attentive to, and adaptive to the other in a conversation", the "immediacy" in the behaviour of the
151   interlocutors, and their degree of "social anxiety".

152   Engagement with an agent is typically referred to as *social engagement*. There is variation in the definition
153   of social engagement. It can be defined as any interaction a human has with either another human being or
154   a robot (Sidner et al. (2003) and Poggi (2007)). Another definition of "social engagement" is provided by
155   Moshkina et al. (2014) as "a core social activity that refers to an individual's behaviour within a social
156   group". The commonality which can be highlighted between both of these two definitions is that social
157   engagement happens in interaction with one another. These definitions remain relatively vague and leave
158   space to encompass a great variety of activities and experimental set-ups with different degrees of socialness.
159   Within the studies reviewed in this article, the activities that could be classified as being more social include
160   storytelling (see for example Szafir and Mutlu (2012)), followed by unstructured conversations and games.
161   While unstructured conversations can encompass many aspects of conversation, generally they are not
162   task-driven.

163   Social engagement very often also includes an affective component, see for example Corrigan et al.
164   (2016); Sohail et al. (2019); Youssef et al. (2019); Biancardi et al. (2019b). The definition of the affective
165   component of engagement often remains vague. It is sometimes related to fun, as is the case in Rehm and
166   Jensen (2015), entertainment, as exemplified in Vázquez et al. (2014). One way in which it is being used is
167   to capture the perception of the inner state of a participant and the value he/she attributes to the interaction,

as in the case of Castellano et al. (2009). Poggi (2007), for example, defined engagement as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and continuing the interaction". Other works highlight the emotion component more predominantly such as Subramainan et al. (2016a), Choi et al. (2012) and Sanghvi et al. (2011). For example, in Sanghvi et al. (2011), user's engagement with the iCat is characterised by an affective and attention component. Similarly also Youssef et al. (2019) rely on several levels of engagement annotation including both the affective as well as the (non-)verbal cue level that includes head rotation and eye-gaze. Finally, Biancardi et al. (2019b) combines affective- with attention- and cognitive engagement components in their detection model. Every person might of course differ to a certain degree in the way that he/she expresses the different forms of engagement.

Engagement with a task is typically referred to as *task engagement*. There is great variety in definitions of task engagement. On the one hand, task engagement is defined as a human involved in a task (Corrigan et al. (2015)). In such a context, the human does not interact with an interface, a robot, an agent to perform the task. Since our focus is on human-agent interaction, we will not consider this case. Rather we will consider task engagement in the context of human-agent interaction where a human and an embodied agent interact together around a task. On the other hand, task engagement can also refer to any kind of human-agent interaction in which behaviours are centered around a task. Examples of such interactions include the one of moving objects, or an object learning experiment where the agent asks participants to identify the name of objects so that it can learn them (Ivaldi et al. (2014)), or mobile robots approaching humans (Ramírez et al. (2016))

In recent years there has also been an increasing amount of work going beyond dyadic to group interactions. This includes work on engagement as well. There are different ways to approach quantifying engagement in a group. Gatica-Perez et al. (2005) defined group interest as "the perceived degree of interest or involvement of the majority of the group". Salam et al. (2017) defined "group engagement" as "the engagement state of two entities in the interaction together with another entity". Oertel et al. (2011) defined group involvement as "a group variable which is calculated as the average of the degree to which individual people in a group are engaged in spontaneous, non-task-directed conversations". Similarly, Salam et al. (2017) defined group engagement as " the engagement state of two entities in the interaction together with another entity". They stress the importance of distinguishing group engagement from other group related constructs such as "cohesion" Carless and De Paola (2000) and "mutual engagement". Goffman (2017), who built on Clark (1996), refers to people within an interaction as belonging to different participation categories. To make this classification, he first distinguishes between participants and non-participants. The group of participants he considers consist of "the speaker", "the current addressee" and "the sideparticipant". The group of non-participants includes the categories of "bystanders" and "overhearers". This highlights that a group is not a simply a set of dyads.

Finally, engagement is also being investigated within the context of long-term human-agent interaction. In such a context Trinh et al. (2018) separate between 3 user categories: Those who "dropout", those who are "moderately engaged" and those who are "highly engaged".

As can be seen in Table 1, there is quite some variation in roles taken on by the virtual agent or robot respectively. No clear distinction becomes apparent between roles taken on by the robot or roles taken on by the agent. However, the more recent papers seem to use more often the role of a conversational partner. At the same time many of the recent papers also have an affective component to their engagement definition. A probable reason for this might be that recent developments have lead to great improvements

211 in multi-modal sensing in general and speech recognition in particular, which makes the implementation of
212 mixed-initiative interactions more feasible.

213   It can be concluded that there are several definitions of engagement. These definitions are not contradictory
214 but rather complementary. They show that engagement is a multi-faceted phenomenon. However, we
215 also notice that the majority of papers reviewed do not directly define engagement or remain vague in its
216 definition. Similarly, papers do not always make a clear distinction between task and social engagement as
217 the scenarios they use often involve both social and task components. For example, a robot as an assistant
218 is a common scenario. However, the role of an assistant comprises both social and task components.

4   SITUATED INTERACTION

219 In the following sections we are referring to engagement as it pertains to interaction in general. In many of
220 the examples given, however, engagement is discussed in the context of conversation. We see and reviewed
221 engagement, however, in general and see conversation as a specific instance of interaction. The fact that
222 conversation as an interaction scenario is represented particularly frequently is an artifact of the papers
223 reviewed.

224   It appears that situations and scenarios do vary considerably across studies. Examples, of such scenarios
225 include museum guides (Pitsch et al. (2009); Salam and Chetouani (2015); Biancardi et al. (2019b)), games
226 (Klotz et al. (2011); Leite et al. (2014); Díaz et al. (2011)), hospitality (Sidner and Lee (2003); Sidner
227 and Dzikovska (2005)), education (Papadopoulos et al. (2016); Leyzberg et al. (2014)), sales (Ishii et al.
228 (2011)), and receptionist/direction giving (Bohus and Horvitz (2014); Michalowski et al. (2006)). With
229 regards to museum guides, the overarching goal in the studies has been on making visitors more interested
230 in the artwork and to provide information on the artwork on demand. Such studies have been done both
231 with virtual agents and with robots. These interactions require a knowledge base of the paintings as well as
232 interaction management with the users. Interactions in general are made more difficult by the surroundings
233 in which the interaction is taking place. In a museum generally the noise level is high, participants are
234 passing by in an uncoordinated manner and the number of interaction partners can vary from one to
235 many. This situation puts certain constraints on human-robot interaction, such as the robustness of speech
236 recognition and turn-taking regulation. However, not all interactions rely on dialogue. Some Human-Robot
237 interactions are text-based or picture based. In addition also the degree of interactivity varies. Some systems
238 are based more on providing information on demand whereas others act more interactively (Bohus and
239 Horvitz (2009b)). Other interactions are in contrast being designed for or already situated in a home
240 environment (for example, supporting health-care, or also in the context of a social companion) Sidner
241 (2016). Researchers have many purposes. Generally in a context like these there is not so much noise, the
242 environment is quieter and the interaction is more focused on a single user or a smaller group of people.
243 However, the interaction is also required to go more into depth and social aspects are more important. Also,
244 memory and variability of interaction become more essential. One further aspect is that of an experiment
245 being situated within a controlled laboratory environment versus an open, uncontrolled environment.

246 ### 4.1   Lab or Real World

247   If we want to go towards more long-term interactions and away from very context dependent interactions,
248 then it is important to gather data and build models of how interactions happen in real-world situations.
249 Questions then arise including how to capture changes of engagement over the course of an interaction,
250 how group size effects engagement and what implications this has on model building. To capture more of
251 these conversational dynamics, Oertel et al. (2013) recorded 5 participants over the course of two days. All
252 interactions were recorded with audio and video and also motion captured. Instead of providing participants

with a task or guiding the interactions, they were simply asked to talk to one another. While this data led to several research papers on investigating engagement, it was also limited in that still several multi-modal cues needed to be annotated manually. Specifically, the engagement label itself but also lower-level cues such as eye-gaze and voice activity still needed to be annotated manually. Oertel et al. (2014) created a corpus which allowed them to study engagement in a setting that fostered free flowing interaction. However, at the same time it was much more controlled in terms of interaction phases. Participants were much more restricted in terms of conversation topic (i.e. their PhD studies) and movement (they were asked to remain seated around a table). Due to this set-up it was however possible to infer gaze and speech activity automatically as well as to optimise for changes in conversational dynamics.

There are more corpora available which try to capture engagement in lab or in the-wild settings to different degrees. Many of them are however not publicly available. Such corpora are then often used in order to predict engagement states of participants, examples include Kim et al. (2018), Kim et al. (2016a), Oertel and Salvi (2013) and Oertel et al. (2011).

In the HRI community, research is going more and more towards bringing robots out of the classical lab environment and evaluating them "in the wild". The notion of what "in the wild" entails often remains unclear. In the following section we are going to highlight differences between in the wild and lab settings and point towards challenges and advantages associated with each scenario. The vast majority of studies reviewed are lab studies. The "in the wild" studies were mainly concerned with long-term interaction or robot child-interactions or both. Examples of child-robot interaction scenarios often concern museum environments. The manner in which the interaction is then realised can vary widely. For instance, Rehm and Jensen (2015) and Siegel et al. (2009) evaluate engagement in the context of a museum. However, in the former case, the agent is a monster agent which ate artworks and the children had to find information about these cultural artefacts. This interaction is quite different to an experimental set-up which is centered around conversations between an agent and a human. Ahmad et al. (2017), for example, carry out a long-term interaction experiment in which children play a game of snakes and ladders with a Nao robot. The children's engagement is later analysed offline. Similar to this experiment, also Moshkina et al. (2014) carried out their experiments in a public space but here the difference was that the authors were interested in short-term interaction. That interaction revolved around story telling and the experiment was more concerned with investigating how a change of different multimodal cues of the agent effects the social engagement of the human.

A middle ground for lab or real-world settings is the school setting as for example used in Zaga et al. (2015), Castellano et al. (2017) and Leite et al. (2014). Students are situated in environments they are more familiar with but most school experiments are still pull-out studies. This means that experiments are typically more controlled than if they were carried out in a public space such as a museum. Castellano et al. (2017) look at the effect of task, social context and their interdependencies in human–robot interaction. Yet, it needs to be noted that engagement is here only evaluated indirectly through an assessment of interaction quality. Castellano et al. (2014), also in a school context, evaluated the effect of robot initiative on the students learning task and the perceived engagement. Similarly, Leite et al. (2014) investigated the effect of an empathetic robot designed for long-term interaction on social presence, engagement and perceived support in children. Similar scenarios are also found in Castellano et al. (2017), Castellano et al. (2014), Leite et al. (2014) and Castellano et al. (2013).

Pitsch et al. (2009) investigated contingent versus non-contingent behaviour strategies and the effect on the engagement of the user. This approach is quite different in terms of the scenario than the more game oriented approaches listed earlier; although, also here the task of the robot was to provide information

297 about paintings. Moshkina et al. (2014) investigate the behaviour of a robot in a public space. Similar to
298 the previous examples, the scenario is concerned with story telling and game playing and is notable for the
299 large number of participants (around 400) involved in the experiment. Alač et al. (2011) use a game-like
300 interaction to focus on the interaction of a robot and the effect of different activities on the engagement of
301 children. An additional scenario in the wild, conducted by Šabanović et al. (2013), involves studying the
302 effect of robots on participant engagement in nursery homes. Finally, in the wild scenarios also include
303 office environments in which agents take on the role of receptionists (Bohus and Horvitz (2009b)) or
304 direction givers (Bohus and Horvitz (2014)).

305 ## 4.2 Annotations

306 The way in which engagement is being measured is crucial for its quantification and generalisability
307 across different studies. However, in a similar manner to defining engagement, conducting annotations
308 of engagement is also a challenging endeavour since there are no generally accepted and established
309 annotation schemes in use. With regards to perception studies, there are two general approaches for
310 annotating engagement. The first relies on first-person annotations of the interaction and the second
311 one on third-party observer annotations. First-party observer annotations are generally provided at the
312 session level. While this approach has the advantage that it captures the perception of the participants
313 in the study, it has the disadvantage that it is not well suited for capturing the conversational dynamics
314 within the session. To capture conversational dynamics within a session, a third-party observer approach
315 using thin-slicing, see for example Ambady and Rosenthal (1993), might be better suited. In a thin-sliced
316 approach, video segments ranging from 5-30 seconds are extracted. These segments of audio and/or video
317 are then being rated by several annotators and inter-rater reliability is calculated. The scale on which
318 engagement is measured also determines how inter-rater reliability measured. There are several ways of
319 conducting the annotation. Some approaches use a binary annotation scheme whereas others use a scalar or
320 ordinal scale annotation schemes. Commonly used inter-rater reliability measures are Cohen's kappa, Fleis'
321 kappa and Krippendorf's alpha. Regarding the generation of engagement, there does not seem to be an
322 equally established trend. Engagement and disengagement are here often associated with the presence or
323 non-presence of an interaction event.

5 PERCEPTION

324 The vast majority of studies reviewed here investigate engagement from a perception system point of view
325 which means that sensors are used as a primary means of input for estimating the degree of engagement
326 or an engagement state. One common task is to estimate a human user's engagement in a conversation
327 or task and to then adapt the robot's behaviour accordingly. The purposes and research questions behind
328 it vary widely which is also linked to the issue of differing definitions of engagement. For example, a
329 case in which engagement is defined by the proximity to a robot will have very different implications for
330 sensors and the perception system than in a case where engagement is defined by subjects' participation
331 in a conversation. Another different approach is the perception of engagement through visual attention
332 mechanisms. Yet again, the perception of signals related to engagement is used to evaluate the impact
333 different tasks have on the unfolding of conversations. A further application for the perceptual detection of
334 engagement is long-term interaction. In order to engage a human for a longer period of time, engagement
335 detection and reactions are becoming increasingly important.

336 ### 5.1 Automatic Prediction of Engagement

337 Several works have addressed the automatic prediction of engagement. In general a distinction can
338 be made between rule-based and machine learning-based approaches for the prediction of engagement.
339 Examples of studies that report rule-based approaches are Brown et al. (2013), Glas et al. (2015), Ishii and

340 Nakano (2008) and Ponsler (2011). There are differences in how rules are implemented. For example, in
341 Ishii and Nakano (2008), rules are based on gaze-transition patterns, whereas in Brown et al. (2013) they
342 are based on the speed of students' responses to a math test. In many studies however, *social engagement*
343 is measured through eye-gaze (see Nakano and Ishii (2010) and Qureshi et al. (2013)), due to the close
344 relationship between *visual attention* and engagement. Ishii et al. (2013a), for example, found that the use
345 of various gaze features provides a good estimate of the user's conversational engagement.

346   In addition to eye gaze, other measures of user attention have been proposed in the literature. Szafir and
347 Mutlu (2012), for example, designed a system that allows a robotic agent to monitor student attention in
348 real-time using measurements from electroencephalography (EEG).

349   As far as machine learning-based methods are concerned, while the proposed methods vary to a great
350 extent, many of the studies rely on Support Vector Machines (SVMs) and use eye-gaze or head-pose as
351 input-features for the engagement prediction. Oertel and Salvi (2013) investigated individual involvement
352 and group engagement. They used different gaze variables to first summarise and explain group actions and
353 then to investigate whether changes in these variables are good predictors of engagement. Moving beyond
354 user attention as a measure of engagement, works by Sanghvi et al. (2011) and Castellano et al. (2014)
355 found that patterns of postural behaviour can be used to accurately predict the engagement of children with
356 a robot during game play and that the latter can also be predicted using information about the children'
357 valence, interest and anticipatory behaviour. In their work, social engagement is modelled as a state
358 consisting of affect and attention components in the context of the interaction. Castellano et al. (2009) also
359 found that an approach that includes both task and social interaction-based features to measure engagement
360 with a robot outperforms those based solely on non-verbal or contextual information. Castellano et al.
361 (2017) showed that game and social context-based features can be used to predict engagement with a robot
362 and that their integration with context-based features encoding their interdependencies leads to higher
363 recognition performances. Kim et al. (2016b) proposed to approach the automatic prediction of engagement
364 using an ordinal learning method and showed that such a method can successfully be used to predict
365 children's engagement using non-verbal features. Foster et al. (2017) used Conditional Random Fields
366 to predict engagement in multi-party HRI using audio-visual data. In their work the task was to estimate
367 engagement of customers for a robot bartender based on the data from audiovisual sensors, which relates
368 to the need for a robot in a dynamic real world environment to infer people's intentions in the scene in
369 order to only attend to those who wish to interact with it. Ishii and Nakano (2010) found that taking into
370 account individual differences of users in gaze transition patterns performs the best in predicting user's
371 conversational engagement. Ishii et al. (2011) then extended their model by adding to gaze transition
372 patterns other gaze parameters such as the occurrence of mutual gaze, gaze duration, distance of eye
373 movement (toward objects of interest in the interaction), and pupil size. They found that considering gaze
374 behaviours in their complexity enhances the performance for predicting user's conversational engagement.

375   In addition to verbal and non-verbal behaviour and context, other modalities and social variables have
376 been identified as important for the automatic prediction of engagement. Choi et al. (2012), for example,
377 found that people's physiological reactions such as heart rate and electrodermal activity can predict the
378 extent to which people will engage affectively or strategically with an agent. Moreover, Salam et al. (2017)
379 found that taking into account personality for the classification of engagement is important. Similarly,
380 Ivaldi et al. (2017) showed that engagement models classically used in human-robot interaction should
381 take into account attitudes and personality traits.

382

5.1.1 Deep Learning Approaches

383 As mentioned in previous paragraphs, machine learning techniques have been widely used for
384 engagement recognition. The success of these techniques heavily depends on both the choice of data
385 representation (input features) and annotation on which they are applied. Most of the input features are
386 domain specific and data representation usually results in a feature engineering phase, as exemplified
387 in Anzalone et al. (2015). The main advantage is the explanatory dimension of the input features. In
388 Leclère et al. (2016), the percentage of time spent face to face or oriented to the task is used to assess
389 face-to-face and task engagement in clinical settings. However, the features are not easily transferable to
390 new tasks, situations and applications. Improving data representation for classifiers is the main objective of
391 representation learning, as described in Bengio (2011). Deep learning is a specific method for achieving
392 representation learning using multiple non-linear transformations. Representation learning based on deep
393 learning is particularly of interest in multimodal processing of human behaviour data by reducing the need
394 of priors on the nature of relations between modalities, the dynamics of non-verbal signals, nature of the
395 task and their impact on the prediction of socio-cognitive states such as engagement.

397 In Rudovic et al. (2019a), a deep learning approach called PPA-net (Personalized Perception of Affect
398 network) is introduced to jointly analyse visual (face and body), audio and physiology data for the prediction
399 of valence, arousal and engagement in autism therapy. The network is designed with three layers: (i) a
400 feature layer, learning representation of each modality, (ii) a context layer, processing of heterogeneous data
401 and expert knowledge, and (iii) an inference layer, predicting the level of arousal, valence and engagement.
402 Feature representation learning is performed by Auto-Encoders (AE), which transform signals to a hidden
403 representation. Interestingly the approach allows one to integrate the correlations among modalities into
404 the representation learning. The context layer aims to augment the feature representation with expert's
405 inputs, which are domain specific (mainly the assessment of children). The last layer is a multitask learning
406 phase, which aims to learn child-specific layers for valence, arousal and engagement estimation. Taken all
407 together, this architecture allows learning correlations between modalities, introducing expert knowledge,
408 personalisation as well as relations between affective states.

409 Another strong motivation for deep learning approaches is learning the dynamics between features. Even
410 with the use of explainable features such as head pose, the relationship between the dynamics of such
411 features and engagement is not always straightforward. Explicitly learning the temporal dynamics between
412 the features as the mapping to engagement could be performed by deep learning approaches. In Hadfield
413 et al. (2018), a Long Short-Term Memory (LSTM) neural network is employed to classify engagement
414 of children to the task using pose data. LSTM are recurrent neural networks able to capture the different
415 dynamics of time series and they have been shown to be efficient in sequence prediction problems. These
416 models have been successfully applied to engagement recognition using head movements in Hadfield et al.
417 (2018) and Lala et al. (2017) and facial expression in Dermouche and Pelachaud (2019a). Temporal models
418 such as LSTM and Gated Recurrent Unit (GRU) are compared to static deep leaning approaches as well as
419 logistic regression. The results show that temporal dynamics as well as the observation window and buffer
420 delay are important factors in the performance of classifiers.

421 All these approaches rely on the availability of engagement annotation. Recently Rudovic et al. (2019b)
422 propose a multimodal active learning approach based on deep reinforcement learning to find the optimal
423 policy for active selection of the user's data. The classification of individual modalities into engagement
424 levels (high/low/medium) is performed by LSTM models followed by fully-connected layers. The output
425 of classifiers are also fed to a Deep Reinforcement Learning agent (Q-function). The agent receives a
426 reward related to its decision: a positive reward is given for correct predictions, and negative rewards is

427 given for incorrect predictions or label requests to human expert. This approach is designed for adapting
428 the engagement prediction model to new tasks, using a minimum number of queries. In addition, as in
429 Rudovic et al. (2019a), the approach also allows multimodal processing and personalisation.

430

5.2 Automatic Prediction of Disengagement

432   While most of the works reported in the literature address the automatic prediction of engagement,
433 examples of systems capable of automatically predicting user disengagement can also be found. Leite
434 et al. (2016), for example, developed an algorithm to predict disengagement in small groups of children
435 interacting with two robot actors playing out interactive narratives around emotional words, using visual
436 and auditory features such as voice activity, smiles, postural and gaze behaviour. Bohus and Horvitz
437 (2014) also investigated disengagement. They used proximity, stability and attention persistence in order
438 to estimate the degree of user disengagement. Conversational hesitation markers were used to estimate
439 whether a participant is still interested in continuing to engage in a conversation.

440   More recently, Youssef et al. (2019) addressed the detection of engagement decrease in users
441 spontaneously interacting with a socially assistive robot in a public space. Recurrent and Deep Neural
442 Networks were used to detect user engagement decrease in real-time based on analysis of user's behaviours
443 such as proxemics, gaze, head motion, facial expressions and speech.

6 GENERATION

444 Engagement aware behaviour generation can be accomplished through a multitude of strategies, which
445 is also exemplified in the papers reviewed in this article. First of all, it can be noted that behaviour
446 generation is dependent on its target audience. Engagement strategies will have to be adapted for children
447 or for people with special needs. Moreover, strategies for behaviour generation are also dependent on
448 the conversational contexts. Humans are generally very good at estimating an interlocutor's level of
449 engagement and reacting appropriately. Agents need to learn more explicitly when to engage and when
450 to disengage from a conversation. Strategies to achieve exactly this can be both based on verbal cues as
451 well as audio-visual ones. Concerning visual cues, a very important process is the establishment of joint
452 attention, either guiding the interlocutor's attention towards an object or indicating its shared focus of
453 attention by reciprocating the focus of attention.

454   Investigating the impact of robot types of autonomy (robot teleoperated by a remote operator vs
455 autonomous robot) on emotional engagement, Choi et al. (2014) reported that participants felt more social
456 presence to teleoperated robots than autonomous robots. Moreover, participants felt more embarrassment
457 when they were interviewed with teleoperated robots than autonomous, for example Baek et al. (2014)
458 found that participants felt more social presence in the company of a person, than a product or a robot. As
459 in Choi et al. (2014), the authors observed that an autonomous robot is not able to easily invoke social
460 emotions. Short et al. (2010) found that participants displayed a greater level of social engagement and
461 made greater attributions of mental state when playing against the robot in the conditions in which it
462 cheated. Moshkina et al. (2014) reported that the more human-like the robot behaves during story-telling,
463 the more social engagement was observed. However, robot's game-playing did not elicit more engagement
464 than other, less social behaviours.

6.1 Adaptation mechanisms

466   The following section provides an overview of studies concerned with strategies used for generating
467 behaviours both adapted to conversational context and user group, especially focusing on the adaptation
468 mechanisms used. During an interaction, interlocutors adapt to each other at various levels. Several

469 computational models have been proposed to decide when the robot should display a particular behaviour
470 or use a specific conversational strategy to call for user's engagement or increase it.

471   Several works report how the timing of agent's behaviour in relation to user's behaviour is important in
472 maintaining the engagement and enhancing user's experience of the interaction. Ishii et al. (2013b)
473 find that the use of probing questions by the engagement-sensitive agent successfully recovers the
474 subject's conversational engagement, changes the gaze behaviours of the participant and elicits more
475 verbal contribution. Sidner et al. (2003) argue that, in robot-user interaction, users are sensitive to the
476 robot's gaze and gestures. They found that a robot's gestures attract the attention of users. The authors
477 also report that users gazing at the object relevant to the conversation at the same time as the robot is
478 a strong sign of user engagement. So detecting user's behaviour in relation to the conversation and to
479 the robot's behaviour is crucial to gain information about how users participate in the interaction. They
480 also note that the rules for driving a robot's engagement maintenance behaviour must be more complex
481 than simply copying users' behaviour. Robins et al. (2005) found that the provision of feedback from the
482 robot in a timely manner was important for the interaction as well as the rhythm, timing of movement
483 and turn-taking in general. Xu et al. (2013) report similar results for multi-party interaction. The authors
484 conducted an experiment where a robot interacted with multiple people at once. There were two conditions:
485 the first involved the robot gazing at its main interlocutor and managing the distribution of turn-taking
486 between interactants, while the second involved the robot gazing and managing turn-taking randomly. The
487 results show that when the robot shows engagement-aware behaviours in the first condition, it significantly
488 improved the effectiveness of communication and positively affected users' experience.

489   While most studies consider gaze behaviours, Cafaro et al. (2016) studied how different interruption types
490 affect the perception of engagement. Interruptions may be cooperative when the interrupter participates to
491 the ongoing conversation by asking for clarification, showing agreement, while they may be disruptive
492 when the interrupter shows disagreement, changes topic of conversation, etc. Cafaro et al. (2016) found
493 that when using a cooperative interruption strategy such as completing the speaker's sentence or asking a
494 clarification question, e.g. to increase affiliation, i.e., liking or friendliness (as opposed to a disruptive one
495 that includes showing disagreement or changing topic of conversation), an interrupter is perceived as more
496 engaged and more involved in the interaction.

497   Other works propose learning approaches for the agent to increase user's engagement. To this aim, the
498 agent learns how to adapt to user's behaviour. Szafir and Mutlu (2012) developed an adaptive robotic
499 agent that employs verbal and non-verbal immediacy cues, such as modulating spoken volume and using
500 gaze, head nodding, and gestures, to regain attention during drops in engagement. The robot acted as
501 an instructor telling stories to students wearing EEG headsets. A trained model to detect reductions in
502 engagement from the EEG data was used to trigger robot's immediacy cues. A robot displaying adaptive
503 behaviours to students' behaviours improved their recall abilities. For female students, their motivation in
504 the education task and rapport with the robot.

505   Keizer et al. (2014) explored social state recognition in multi-party HRI, with a specific focus on
506 building machine learning methods to determine whether a user within a group is seeking to engage with a
507 robot bartender using a combination of multimodal features. Markov Decision Processes (MDPs) were
508 employed to generate socially appropriate behaviour by the bartender robot based on each individual user's
509 engagement.

510   Pelachaud and colleagues have developed several adaptation mechanisms to control a virtual agent
511 whose aim is to maintain user's engagement during the interaction. The adaptation mechanisms work at

512 three different levels: the nonverbal behaviour, conversational strategy and signal levels. Each of these
513 mechanisms have been implemented in the same agent architecture but affect different modules. Adaptation
514 of the conversational strategies is done within the *dialog module* described in Biancardi et al. (2019b). The
515 adaptation at the non-verbal level is done in the *behaviour generation module* of the Greta platform, as
516 described by Biancardi et al. (2019c). Finally, the adaptation at the signal level is done within the *animation*
517 *module* found in Dermouche and Pelachaud (2019b). During an interaction with a user, the agent will
518 optimise each of the adaptation mechanisms by relying on either reinforcement learning or LSTM methods.
519 Biancardi et al. (2019b) conducted an experiment in a science museum where the agent played the role of a
520 museum guide to validate their adaptation models. The agent that adapted its behaviour to maximise user's
521 engagement was perceived as warm by participants, but they did not find any effect of agent's adaptation
522 on users' evaluation of their experience of the interaction. As noted in Dermouche and Pelachaud (2019a),
523 engagement was defined by user's behaviours that included gaze directions, facial expressions and posture
524 shifts.

525   Bickmore et al. (2011) found that the use of relational behaviour lead to significantly greater engagement
526 by museum visitors. In this study, that had 1607 visitors participating, engagement was measured by
527 session length, number of sessions, and self-reported attitude, as well as learning gains, measured by a
528 knowledge test.

529   Other works have explored personalised tutoring from the perspective of affective policy learning: for
530 example, affect-related states such as engagement have been used by Gordon et al. (2016) to build reward
531 signals in reinforcement learning (RL) frameworks to select motivational strategies. Gao et al. (2018)
532 developed an RL framework for robot personalisation that allows a robot to select verbal supportive
533 behaviours to maximise the user's task progress and engagement (i.e., positive reactions towards the robot)
534 in a learning scenario where a Pepper robot acts as a tutor and helps people to learn how to solve grid-based
535 logic puzzles.

536   Some papers explicitly focus on engagement towards a task. For example, Zaga et al. (2015) found that
537 students engaged and focused more on a task (puzzle) when the robot acts as a peer than as a tutor. Brown
538 et al. (2013) reported that engaging with the robot during a computer-based math test showed that, while
539 various forms of behavioural strategies increase test performance, combinations of verbal cues result in a
540 slightly better outcome. Ivaldi et al. (2014) found that whether the robot or human initiates the learning
541 task makes a difference on the pace of the interaction and the reaction to attention cues. When the robot is
542 the initiator of the learning task, the pace of interaction is higher and the reaction to attention cues faster.
543 Whether the robot initiates the interaction does not affect the perceived engagement.

544   Some studies focused on children interacting with robots. Brown and Howard (2013) monitored students'
545 engagement levels while conducting math exercises in the presence of a robot using interaction features
546 such as speed and validity of submitted answers. When student disengagement was detected, for example
547 when there was inactivity for too long or the student was not challenged enough, the robot would employ
548 verbal and non-verbal behaviours that were found to reduce children's boredom during the education
549 task (Brown et al. (2013)). These behaviours could be a combination of socially supportive utterances,
550 backchannels, gaze contact, gestures, and head movements. Leite et al. (2016) developed an algorithm to
551 monitor engagement in small groups of children interacting with two robot actors and trigger disengagement
552 repair interventions when necessary. With the help of elementary school teachers, repair strategies were
553 designed. They include the robots addressing the whole group and making generic comments that imply
554 responsibility of all participants, looking at each of the children in the group, and then generating a verbal
555 comment without targeting any specific child, or directly addressing the child with the highest level of

556   disengagement. They found that participants who received targeted interventions had higher story recall
557   and emotional understanding, and their valence increased over time after the interventions.

7   LONG TERM INTERACTION

558   Longitudinal interactions or "long-term interactions" in HRI are defined by several researchers as a set of
559   interactions over several sessions. For example, Leite et al. (2013) address the question on what defines
560   a long-term interaction with a robotic agent. In their work, they state that the main aspects to define a
561   long-term interaction are based on the number of interaction sessions and the duration of each session.
562   Specifically, they suggest that what constitutes long-term interaction is the point in time when the novelty
563   effect wears off.

564   Adopting the definition by Leite et al. (2013) on long-term interaction, we could find only four papers
565   explicitly investigating long-term interactions and engagement: Ahmad et al. (2017), Leite et al. (2014),
566   Díaz et al. (2011) and Bickmore et al. (2010).

567   Ahmad et al. (2017) presented a study with children to investigate adaptive capabilities of a robot that can
568   sustain a long-term social engagement when interacting with children. In their work, they designed a study
569   that has three adaptive conditions that include game-based adaptations, emotion-based adaptation, and
570   memory-based adaptation. Their study had 23 school participants (aged 10-12 years) that were randomly
571   assigned to each of the three conditions. Participants were asked to do three recorded sessions (10 minutes
572   for each session) over a period of 10 days. The authors, thereafter, conducted video based analysis to
573   investigate the participant's facial expression, gaze, verbal interaction and gestures. Their analysis revealed
574   that the emotional-based adaptive robot maintained a longer interaction compared with the game-based and
575   memory based adaptive robots.

576   Leite et al. (2014) presented an empathic model that is designed to allow for a long-term interaction. In
577   their work, they present a study with 16 elementary school children (8-9 years) interacting with a robot in a
578   chess game scenario. The aim of their study was to investigate children's perception of social presence,
579   engagement and support towards the robot over time. Their study extended over five weeks in which each
580   participant had one session per week (average of 20 minutes per session). The authors collected data
581   from open ended interviews, questionnaires, video recordings, affect sensors, and game play logs. Their
582   analysis revealed that in long term child-robot interactions, a robot endowed with an empathetic model was
583   perceived to support children in a similar manner to the support received from their peers. Furthermore,
584   children found the emotion-based support to be their preferred supportive behaviour, echoing the results
585   presented by Ahmad et al. (2017).

586   Díaz et al. (2011) aimed to build a robot capable of maintaining the engagement of children. The authors
587   presented a study to investigate the social bonds associated with child-robot interaction in order to design
588   for long-term interaction for hospitalised children. In their study, they had two phases, the first was a field
589   study using four robots at a school with 49 children (11-12 years). The second phase had 7 children from
590   the 49 to contribute in a lab study with one of the robots two months later. Data collected from both phases
591   include video, questionnaires, observations, interviews and interaction choices. Their analysis reveals
592   that appearance and performance aspects are important design considerations for long term interaction, in
593   particular, they shape the expectations of children towards interactive behaviours and social processes.

594   Bickmore et al. (2010), on the other hand, presented two studies on maintaining engagement in a long-
595   term interactions with virtual agents. In the first study, the authors investigated the effect of the agent's
596   dialogue repetitiveness behaviour on retention and adherence for elderly persons. Elderly participants were

597 randomly assigned to each of the two conditions (variable dialogue and non-variable repeated dialogue).
598 Participants interacted with the virtual agent 102.32 days on average. The results of the first study revealed
599 a positive effect on the variable dialogue behaviour towards long term engagement. The second study
600 presented by Bickmore et al. (2010), looked at the effect of using personal human back stories, i.e.
601 first person compared to third-person narrative dialogue. In the second study participants were assigned
602 randomly to each of the two conditions. Participants conducted the study over different durations (with an
603 average of 28.8 days). The results revealed that participants rated the first-person dialogue of the virtual
604 agent to be more enjoyable and usable, thus leading to higher engagement.

605   Overall, the presented HRI studies highlight that long term interactions require the support of emotional
606 robotic characteristics, both at the perception level and at the generation level. In addition, adaptive
607 emotional based engagement helps to maintain a long-term interaction.

8   CHILDREN

608 Several studies have investigated engagement in child-robot or child-agent interaction. The majority of them
609 address automatic prediction of engagement and adaptation mechanisms in educational and edutainment
610 scenarios.

611   In Sanghvi et al. (2011), Castellano et al. (2014) and Castellano et al. (2009), Castellano and colleagues
612 proposed computational approaches for the automatic prediction of engagement in children learning to
613 play chess with an iCat robot in a classroom environment with primary school children. They investigated
614 the role that different behavioural and contextual features play in the automatic prediction of engagement.
615 Their work shows that children's affective expressions emerging in a chess play scenario are very subtle and
616 that performance increases for models that include a combination of social interaction and context-based
617 features. Kim et al. (2016a) showed that levels of engagement can be characterised by relative levels
618 between children. Another example is the work by Rudovic et al. (2019a), who addressed the automatic
619 prediction of engagement from videos in preschool children using deep reinforcement learning.

620   When it comes to adaptation mechanisms, Procter et al. (2016) investigated improving conversational
621 engagement through data-driven agent behaviour modification, by adapting to a variety of different student
622 data sources. Szafir and Mutlu (2012) reported that a robot that increases behavioural cues during passages
623 of low student engagement to regain student's attention improved student's recall abilities. Approaches
624 for the automatic prediction of children's disengagement while interacting with a robot in an educational
625 context and disengagement repair strategies have also proposed by Brown and Howard (2013) and Leite
626 et al. (2016).

627   Other works explored the interrelationships between affect, empathy, synchrony and engagement. Hall
628 et al. (2006), for example, found that children are able to recognise and interpret affect in synthetic
629 characters and are empathetically engaged with them in specific scenarios. Chaspari et al. (2015) reported
630 that verbal synchrony between children during game play increases as they become more engaged in
631 a speech-controlled robot-companion game. They showed an interaction with the children's level of
632 engagement: more engaged pairs show higher synchrony in word rate, speech loudness and fundamental
633 frequency. Castellano et al. (2013) showed that a robot's ability to perceive and adapt to children's affect
634 has an effect on their perception of the robot in that children perceive the robot as more engaging and
635 helpful.

636   Finally, a number of studies investigated engagement perception and generation in long-term interaction.
637 Leite et al. (2014) developed an empathic model for child-robot interaction and found that robot empathic

638 behaviour had a positive impact on long-term interaction between children and the robot. They found
639 that ratings of social presence, engagement, help and self-validation remained similar after five weeks
640 of interaction with the robot. Ahmad et al. (2017) proposed adaptation mechanisms to enable a robot to
641 sustain a long-term social engagement when interacting with school children over a period of ten days. In a
642 different context, Díaz et al. (2011) investigated the social bonds emerging in child-robot interaction in
643 order to design for long-term interaction for hospitalised children.

644 In summary, this literature survey shows that engagement has been extensively studied in the context
645 of child-robot interaction. It shows that models for engagement prediction and generation via adaptation
646 mechanisms need to be tailored to the specific end-users i.e. children. This represents both a challenge
647 and an opportunity, as it highlights the need to focus on the development and the study of the effects of
648 personalised technologies, if long-term interactions with robots are to be achieved in the future.

9 OPEN QUESTIONS

649 Even though many models of engagement have been proposed, there are still important open questions that
650 ought to be addressed. We list a few in this section.

651 **Multi-party interaction.** Previous computational models of engagement have focused mainly on dyadic
652 situations. Few models have been proposed for group interaction. Interaction in a group can be very
653 complex. Not everyone in the group may be involved in a task, or participate to the discussion with equal
654 interest. One can address the whole group, a specific person in the group, there may be several sub-groups,
655 etc. Often, the participants in a group conversation are clustered into three main classes: speaker, hearer and
656 over-hearer. In HCI there may be multiple humans and/or multiple robots. It remains a challenge to define
657 engagement in groups given the multiple conformations that a group may have and given the variety of
658 parameters (role, position, relation to name a very few) group members may take. Measuring engagement
659 in multi-party may require considering more features than in dyads to capture the engagement of each
660 individual in the group, or of sub-groups composing the group. Engagement in multi-party may involve
661 more than involvement and emotion component, such as degree of cohesion and collectivity.

662 **Dynamics of engagement process.** Another challenge in defining engagement is related to the dynamism
663 of the process of engagement. Engagement is not a static variable. It is a process that dynamically
664 evolves between (two of more) members of an interaction. They may go through different phases of
665 engagement, ranging from disengaged to fully engaged. Most existing works have focused on detecting
666 when participants are fully engaged; very few looked at disengagement (see Leite et al. (2016)'s work for
667 group disengagement). The question of characterizing the different phases of engagement and defining
668 computational model for their detection is still an open question.

669 **Long-term versus short-term engagement.** A further open question remains on whether long-term
670 engagement and short-term engagement are referring to the same underlying concept or not. There
671 are some obvious differences in the way engagement is conceptualised depending on the duration of
672 interaction. While turn-taking and visual attention appear to be of more importance in shorter-term
673 interaction, variability of generation and social bonds become more prominent in studies related to longer-
674 term interaction. It appears to be logical that short and long-term engagement are related and that there is
675 an intricate interplay between the two. However, we are not aware of any studies that have investigated
676 their relationship so far.

677 **Task versus social engagement.** In this article we discussed studies around task and studies around
678 social-engagement. Most of the studies having haven been carried out so far appear to be in the realm

679 of social engagement followed by studies whose scenarios require a combination between the two. This
680 is very likely also an artifact of currently available sensor technology and thereof resulting limitation in
681 scenario design. There is a high probability, however, that the current developments in sensor technologies
682 will enable a much wider range of scenarios in the near future. A likely application domain appear to be in
683 the factories-of-the future where human-agent collaboration around a task will be in the centre of attention.
684 Another likely application domain appears to be in education where currently a lot of emphasis is being
685 given to bringing robots to the classroom. This will necessitate a stronger focus on the conceptualisation of
686 task engagement and in depth analysis of how social and task engagement are related to one another in a
687 wider range of application scenarios and longer-duration interactions.

688 **Data annotation.** As far as the implementation of systems for the automatic recognition of engagement
689 in human-agent and human-robot interaction is concerned, an open question is how to annotate data
690 corpora to train prediction models. Specifically, identifying an appropriate ground truth remains a challenge.
691 Trends in the automatic affect recognition community have pointed to the need to move towards the
692 automatic prediction of continuous affect (Gunes and Schuller (2013)). Initial work in the direction of
693 annotating continuous signals of engagement or engagement-related variables in videos of human-robot
694 interactions has been conducted (Corrigan et al. (2016)). However, open questions remain about the role of
695 engagement-related variables (e.g. user attention towards robot or towards task) and their interrelationships
696 in task-based interactions, where user and robot (or agent) jointly work on a task, and how this affects the
697 annotation process. Moreover, inter-annotator agreement also remains a challenge for research in this area,
698 due to the multifaceted nature of engagement and its definitions as a process or a state.

699 **On-line adaptation.** Engagement can be considered as a rich cue that could be then employed to adapt
700 or train machine learning and decision-making models. However, developing computational models that
701 are able to adapt on-line to non-verbal cues is a current challenge in machine learning. The few attempts
702 to exploit non-verbal cues for adaptation and learning require the prior definition of a communication
703 protocol (i.e. meaning of non-verbal cues) between the human and the machine (Broekens and Chetouani
704 (2019)). Training adaptive machine learning with non-verbal cues are facilitated by the prior categorization
705 of limited discrete signals such as pointing or stop hand gesture (Najar et al. (2016)). The current on-line
706 adaptive models of engagement exploit a similar approach. In (Khamassi et al. (2018)), the authors develop
707 an on-line adaptation model to changes in human engagement by considering head pose as an estimator of
708 engagement. However, as discussed in this paper, engagement is a complex construct for which adaptation
709 to it will require the analysis of multimodal cues during a longer period of time.

710 **Context specificity and scalability.** Many existing human(s)-robot(s) interaction models have been
711 designed for a specific scenario where many of the variables defining an interaction are pre-defined such
712 as the role of the robot, the task to be performed, the interaction setting, the culture of the participants,
713 even the type of the robot or virtual agent. The models are very context-specific. It is not clear if an
714 engagement model that has been framed so specifically can be applied to any other context. The question
715 of scalability is raised. It is not clear whether a model for a given context may be transferred to another one,
716 or should a new model be drawn. In addition, engagement computational models are usually learned for a
717 specific context and task with a given agent either virtual or physical. How to generalize or transfer such
718 engagement models to other situations is an open-question which in turn not only leads to the problems of
719 defining engagement, data collection and annotation, but also how humans engage in complex interactions
720 with those artefacts. The embodiment of the agent plays a critical role in the nature and quality of those
721 interactions.

722 **Complex real world scenarios.** A further open question is how to continue to extend engagement
723 research to agents capable of handling more complex, real-world scenarios, which require a blend of social,
724 task and environment engagement. These may be multi-party, multi-task interactions and may involve
725 the agent moving or manipulating the environment, thus requiring it to maintain a greater knowledge of
726 aspects that are often considered outside the context of the interaction in current scenarios. Fundamentally,
727 engagement has relationships to attentive processes and states. As the complexity of interactions increases,
728 more robust foundations may be needed to address these issues. For example, computational visual attention
729 frameworks represent a generic way of resolving the allocation of processing resources across multiple
730 social stimuli, tasks and unexpected encounters in the greater environment in a continuous and flexible
731 manner. Currently in many research studies, the potential foci of attention of the agent are restricted to
732 at most two aspects: the task and the human interactor. A more detailed approach is desirable if more
733 realistic scenarios are to be considered, especially those that are to be robust to the complex, uncertain,
734 dynamic environments in the real world in which important interruptions can take place at any time. For
735 example, one can consider the difficulty of modelling a mobile social agent that is to move down a street in
736 a formation of people. Such an agent would need to attend to unexpected obstacles in order to safely avoid
737 them and maintain its formation while engaging in a socially appropriate manner with the others in the
738 group. These activities would likely lead to many differences in behaviour (and in behaviours the agent
739 would be expected to produce) when compared to those observed in contemporary scenarios in which the
740 agent is often static in the environment and facing the interlocutor when the interaction commences.

10 CONCLUSION

741 This article reviewed studies on engagement within the area of human-agent interaction. It can be concluded
742 that there exists a wide range of definitions. In general, distinctions can be made between studies that treat
743 engagement as a process variable and studies that treat engagement as a state variable. Also the emphasis
744 on task or social aspects of engagement varies widely. In the vast majority of cases the distinction is not
745 made transparently.

746 A distinction can be made between studies focusing on the (automatic) perception of engagement
747 behaviours and those that focus on the generation of engagement relevant behaviours. Studies reviewed
748 show that both adapting to a target group but also to conversational context is essential. In this article
749 we provide examples of adaptation mechanisms used. Examples of such mechanisms include the use
750 of probing questions or adapting in terms of verbal and non-verbal behaviour to the user. While most
751 approaches are rule-driven, there are also some approaches that use machine learning for adaptation,
752 including reinforcement learning and social-state recognition. While several adaptation approaches have
753 been explored, they vary widely in approach and scenario chosen. Due to a lack of benchmarks, a more
754 detailed comparison at this stage does not appear to be feasible.

755 In terms of the automatic perception of engagement, studies can be divided into two main groups. Those
756 studies which use rule based approaches and those studies which use machine learning based approaches.
757 While the vast majority of studies that are concerned with the automatic prediction of engagement still
758 use traditional machine learning techniques and are mainly SVM-based, there are also a number of
759 recent studies that use deep learning based approaches. Finally, instead of focusing on detecting binary
760 engagement, or different degrees of engagement, there are also a number of studies that focus on detecting
761 disengagement instead.

762 Regarding robot-child interaction the following conclusions can be drawn. First of all, all studies focus
763 on different aspects of the interaction. These aspects include for example the robot's reaction towards the

764 children's engagement state. This can, but does not necessarily have to be linked to their performance, for
765 example in the context of education. It also seems important that the models for engagement prediction and
766 generation via adaptation mechanisms need to be specifically tailored to children. However, there is still a
767 need to explore the aforementioned point more extensively for definite conclusions. This represents both a
768 challenge and an opportunity, as it highlights the need to focus on the development and the study of the
769 effects of personalised technologies, if long-term interactions with robots are to be achieved in the future.
770 In the studies reviewed no direct comparisons are made to adult-robot interaction, rather the reaction of the
771 children towards the robot were in the center of attention.

772 Regarding engagement designed for long-term interaction, not many studies have been carried out yet.
773 There is also no clear-cut definition of what constitutes long-term interaction although Leite et al. (2013)
774 suggest that it is the point in time when the novelty effect wears off. Emotional adaptation, appearance
775 and performance, variable dialogue and first-person narrative all appear to be contributing positively to
776 long-term interaction.

777 More and more work is also concerned with "in-the-wild" studies in contrast to lab studies. Challenges
778 associated with "in-the-wild" studies are that it is much harder to control the interaction. For example
779 noise and suboptimal light sources can interfere with the sensors; the context of the interaction may vary
780 a lot as the number of participants interacting with the robot. The papers reviewed in this article address
781 these challenges in different ways. Some papers focus on collecting corpora which portray non-task-
782 directed interaction to best model engagement dynamics, whereas other papers directly focus on creating
783 an interaction scenario and test it "in-the-wild". Common scenarios appear to be information-giving, story-
784 telling and game-like interactions. A middle ground between the lab and completely "in-the-wild" location
785 appears to be a pull-out-study in a school setting. Students remain situated in a familiar environment, yet
786 noise level and number of participants etc. can be controlled by the experimenter more easily.

787 In summary, this review covers a broad range of studies on engagement in human-agent and human-robot
788 interaction. To the best of our knowledge, it is the first review on engagement research that reports on how
789 the human-agent and human-robot interaction communities have addressed issues and challenges relating
790 to engagement definitions and implementations in different interaction settings, engagement annotation,
791 and automatic engagement prediction and generation in adaptive human-agent interactions. The picture that
792 emerges is one of engagement as a highly complex phenomenon that permeates human-agent interaction
793 and determines its success over sustained periods of time. We review open questions and challenges for
794 the community, offering the reader a starting point for making new, interesting research contributions in a
795 research area that is still growing.

CONFLICT OF INTEREST STATEMENT

796 The authors declare that the research was conducted in the absence of any commercial or financial
797 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

798 CO wrote the initial draft of this article for which she carried out data analysis and created an initial structure.
799 Together with MO she created the annotation scheme and went on to carry out most of the annotations. She
800 was further involved in research discussions on paper improvements and provided feedback GC, CPel and
801 CPet have initiated discussions on an overview paper on engagement in human-agent interaction, provided
802 feedback on the paper's structure and written sections of the paper. CPel and CPet also assisted in the
803 paper annotation process. JN helped with the annotation of the data and was involved in discussions on

804 structure and content. MO contributed to the discussions and the development of the annotation scheme.
805 He conducted parts of the data annotations and data analysis, written sections of the article, gave feedback
806 on the article's overall structure and revised the article. MC contributed to the structure, data analysis,
807 writing and discussions.

FUNDING

REFERENCES

810 Ahmad, M. I., Mubin, O., and Orlando, J. (2017). Adaptive social robot for sustaining social engagement
811   during long-term children–robot interaction. International Journal of Human–Computer Interaction 33,
812   943–962. doi:10.1080/10447318.2017.1300750
813 Alač, M., Movellan, J., and Tanaka, F. (2011). When a robot is social: Spatial arrangements and
814   multimodal semiotic engagement in the practice of social robotics. Social Studies of Science 41,
815   893–926. doi:10.1177/0306312711420565
816 Alami, R., Clodic, A., Montreuil, V., Sisbot, E. A., and Chatila, R. (2005). Task planning for human-robot
817   interaction. In Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence:
818   Innovative Context-Aware Services: Usages and Technologies (ACM), 81–85. doi:10.1145/1107548.
819   1107574
820 Ambady, N. and Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices
821   of nonverbal behavior and physical attractiveness. Journal of personality and social psychology 64,
822   431–441. doi:10.1037/0022-3514.64.3.431
823 Anzalone, S. M., Boucenna, S., Ivaldi, S., and Chetouani, M. (2015). Evaluating the engagement with
824   social robots. International Journal of Social Robotics 7, 465–478. doi:10.1007/s12369-015-0298-7
825 Baek, C., Choi, J. J., and Kwak, S. S. (2014). Can you touch me?: the impact of physical contact
826   on emotional engagement with a robot. In Proceedings of the second International Conference on
827   Human-Agent Interaction (ACM), 149–152. doi:10.1145/2658861.2658909
828 Bengio, Y. (2011). Deep learning of representations for unsupervised and transfer learning. In Proceedings
829   of the 2011 International Conference on Unsupervised and Transfer Learning Workshop - Volume 27,
830   eds. I. Guyon, G. Dror, V. Lemaire, G. Taylor, and D. Silver (JMLR.org), 17–37
831 Biancardi, B., Mancini, M., Lerner, P., and Pelachaud, C. (2019a). Managing an agent's self-presentational
832   strategies during an interaction (Frontiers), vol. 6, 93
833 Biancardi, B., Mancini, M., Lerner, P., and Pelachaud, C. (2019b). Managing an agent's self-presentational
834   strategies during an interaction. Frontiers in Robotics and AI 6, 93. doi:10.3389/frobt.2019.00093
835 Biancardi, B., Wang, C., Mancini, M., Cafaro, A., Chanel, G., and Pelachaud, C. (2019c). A computational
836   model for managing impressions of an embodied conversational agent in real-time. In Proceedings
837   of the 8th International Conference on Affective Computing and Intelligent Interaction (IEEE). doi:10.
838   1109/ACII.2019.8925495
839 Bickmore, T., Pfeifer, L., and Schulman, D. (2011). Relational agents improve engagement and learning
840   in science museum visitors. In International Conference on Intelligent Virtual Agents, eds. H. H.
841   Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson (Springer Berlin Heidelberg), 55–67. doi:10.
842   1007/978-3-642-23974-8_7
843 Bickmore, T., Schulman, D., and Yin, L. (2010). Maintening engagement in long-term interventions with
844   relational agents. Applied Artificial Intelligence 24, 648–666. doi:10.1080/08839514.2010.492259

845 Bohus, D. and Horvitz, E. (2009a). Learning to predict engagement with a spoken dialog system in
846   open-world settings. In Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of
847   the Special Interest Group on Discourse and Dialogue (ACL), 244–252

848 Bohus, D. and Horvitz, E. (2009b). Models for multiparty engagement in open-world dialog. In Proceedings
849   of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse
850   and Dialogue (ACL), 225–234

851 Bohus, D. and Horvitz, E. (2014). Managing human-robot engagement with forecasts and... um...
852   hesitations. In Proceedings of the 16th International Conference on Multimodal Interaction (ACM),
853   2–9. doi:10.1145/2663204.2663241

854 Broekens, J. and Chetouani, M. (2019). Towards transparent robot learning through TDRL-based emotional
855   expressions. IEEE Transactions on Affective Computing , 1–1doi:10.1109/TAFFC.2019.2893348

856 Brown, L. and Howard, A. M. (2013). Engaging children in math education using a socially interactive
857   humanoid robot. In Proceedings of the 13th IEEE-RAS International Conference on Humanoid Robots
858   (IEEE), 183–188. doi:10.1109/HUMANOIDS.2013.7029974

859 Brown, L., Kerwin, R., and Howard, A. M. (2013). Applying behavioral strategies for student engagement
860   using a robotic educational agent. In Proceedings of the 2013 IEEE International Conference on
861   Systems, Man, and Cybernetics (IEEE), 4360–4365. doi:10.1109/SMC.2013.744

862 Cafaro, A., Glas, N., and Pelachaud, C. (2016). The effects of interrupting behavior on interpersonal
863   attitude and engagement in dyadic interactions. In Proceedings of the 2016 International Conference
864   on Autonomous Agents & Multi-Agent Systems (International Foundation for Autonomous Agents and
865   Multi-Agent Systems), 911–920

866 Carless, S. A. and De Paola, C. (2000). The measurement of cohesion in work teams. Small group research
867   31, 71–88

868 Castellano, G., Leite, I., and Paiva, A. (2017). Detecting perceived quality of interaction with a robot using
869   contextual features. Autonomous Robots 41, 1245–1261. doi:10.1007/s10514-016-9592-y

870 Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. W. (2013). Multimodal
871   affect modeling and recognition for empathic robot companions. International Journal of Humanoid
872   Robotics 10, 1–=23. doi:10.1142/S0219843613500102

873 Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. W. (2014). Context-sensitive
874   affect recognition for a robotic game companion. ACM Transactions on Interactive Intelligent Systems
875   (TiiS) 4, 1–25. doi:10.1145/2622615

876 Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). Detecting user engagement
877   with a robot companion using task and social interaction-based features. In Proceedings of the 2009
878   International Conference on Multimodal Interfaces (ACM), 119–126. doi:10.1145/1647314.1647336

879 Celiktutan, O., Sariyanidi, E., and Gunes, H. (2018). Computational analysis of affect, personality, and
880   engagement in human–robot interactions. In Computer Vision for Assistive Healthcare (Academic
881   Press). 283–318. doi:10.1016/B978-0-12-813445-0.00010-1

882 Celiktutan, O., Skordos, E., and Gunes, H. (2017). Multimodal human-human-robot interactions (mhhri)
883   dataset for studying personality and engagement. IEEE Transactions on Affective Computing 10,
884   484–497. doi:10.1109/TAFFC.2017.2737019

885 Chaspari, T., Al Moubayed, S., and Fain Lehman, J. (2015). Exploring children's verbal and
886   acoustic synchrony: Towards promoting engagement in speech-controlled robot-companion games.
887   In Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence (ACM),
888   21–24. doi:10.1145/2823513.2823518

889 Choi, A., Melo, C. D., Woo, W., and Gratch, J. (2012). Affective engagement to emotional facial
890   expressions of embodied social agents in a decision-making game. Computer Animation and Virtual
891   Worlds 23, 331–342. doi:10.1002/cav.1458

892 Choi, J. J., Kim, Y., and Kwak, S. S. (2014). Are you embarrassed?: The impact of robot types on
893   emotional engagement with a robot. In Proceedings of the 2014 ACM/IEEE International Conference
894   on Human-Robot Interaction (ACM), 138–139. doi:10.1145/2559636.2559798

895 Clark, H. H. (1996). Using language (Cambridge university press)

896 Coker, D. A. and Burgoon, J. (1987). The nature of conversational involvement and nonverbal encoding
897   patterns. Human Communication Research 13, 463–494

898 Conrad, F. G., Schober, M. F., Jans, M., Orlowski, R. A., Nielsen, D., and Levenstein, R. (2015).
899   Comprehension and engagement in survey interviews with virtual agents. Frontiers in Psychology 6,
900   1578. doi:10.3389/fpsyg.2015.01578

901 Corrigan, L. J., Basedow, C., Küster, D., Kappas, A., Peters, C., and Castellano, G. (2015). Perception
902   matters! engagement in task orientated social robotics. In Proceedings of the 24th IEEE International
903   Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 375–380. doi:10.
904   1109/ROMAN.2015.7333665

905 Corrigan, L. J., Peters, C., Castellano, G., Papadopoulos, F., Jones, A., Bhargava, S., et al. (2013). Social-
906   task engagement: Striking a balance between the robot and the task. In Embodied Commun. Goals
907   Intentions Workshop ICSR. vol. 13, 1–7

908 Corrigan, L. J., Peters, C., Küster, D., and Castellano, G. (2016). Engagement perception and generation
909   for social robots and virtual agents. In Toward Robotic Socially Believable Behaving Systems - Volume
910   I (Springer International Publishing). 29–51. doi:10.1007/978-3-319-31056-5_4

911 Dermouche, S. and Pelachaud, C. (2019a). Engagement modeling in dyadic interaction. In Proceedings of
912   the 2019 International Conference on Multimodal Interaction (ACM), 440–445. doi:10.1145/3340555.
913   3353765

914 Dermouche, S. and Pelachaud, C. (2019b). Generative model of agent's behaviors in human-agent
915   interaction. In 2019 International Conference on Multimodal Interaction. 375–384

916 Díaz, M., Nuño, N., Saez-Pons, J., Pardo, D. E., and Angulo, C. (2011). Building up child-robot
917   relationship for therapeutic purposes: From initial attraction towards long-term social engagement. In
918   Face and Gesture FG 2011 (IEEE), 927–932. doi:10.1109/FG.2011.5771375

919 Drejing, K., Thill, S., and Hemeren, P. (2015). Engagement: A traceable motivational concept in human-
920   robot interaction. In Proceedings of the 2015 International Conference on Affective Computing and
921   Intelligent Interaction (ACII) (IEEE), 956–961. doi:10.1109/ACII.2015.7344690

922 Foster, M., Gaschler, A., and Giuliani, M. (2017). Automatically classifying user engagement for
923   dynamic multi-party human-robot interaction. International Journal of Social Robotics 9, 659–674.
924   doi:10.1007/s12369-017-0414-y

925 Foster, M. E., Gaschler, A., and Giuliani, M. (2013). How can i help you?: Comparing engagement
926   classification strategies for a robot bartender. In Proceedings of the 15th ACM on International
927   Conference on Multimodal Interaction (ACM), 255–262. doi:10.1145/2522848.2522879

928 Gao, Y., Barendregt, W., Obaid, M., and Castellano, G. (2018). When robot personalisation does not help:
929   Insights from a robot-supported learning study. In Proceedings of IEEE International Symposium on
930   Robot and Human Interactive Communication (IEEE), 705–712. doi:10.1109/ROMAN.2018.8525832

931 Gatica-Perez, D., McCowan, L., Zhang, D., and Bengio, S. (2005). Detecting group interest-level in
932   meetings. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal
933   Processing (IEEE), vol. 1, I/489–I/492. doi:10.1109/ICASSP.2005.1415157

934 Glas, N., Prepin, K., and Pelachaud, C. (2015). Engagement driven topic selection for an information-giving
935   agent. SEMDIAL 2015 goDIAL , 48

936 Gockley, R. and Matarić, M. J. (2006). Encouraging physical therapy compliance with a hands-off mobile
937   robot. In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction
938   (ACM), 150–155. doi:10.1145/1121241.1121268

939 Goffman, E. (2017). Interaction ritual: Essays in face-to-face behavior (Routledge)

940 Gordon, G., Spaulding, S., Westlund, J. K., Lee, J., Plummer, L., Martinez, M., et al. (2016). Affective
941   personalization of a social robot tutor for children's second language skills. In Proceedings of the
942   Thirtieth AAAI Conference on Artificial Intelligence (AAAI Press), 3951–3957

943 Gunes, H. and Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current
944   trends and future directions. Image and Vision Computing 31, 120–136. doi:10.1016/j.imavis.2012.06.
945   016

946 Hadfield, J., Chalvatzaki, G., Koutras, P., Khamassi, M., Tzafestas, C. S., and Maragos, P. (2018). A deep
947   learning approach for multi-view engagement estimation of children in a child-robot joint attention task.
948   CoRR abs/1812.00253

949 Hall, J., Tritton, T., Rowe, A., Pipe, A., Melhuish, C., and Leonards, U. (2014). Perception of own and
950   robot engagement in human–robot interactions and their dependence on robotics knowledge. Robotics
951   and Autonomous Systems 62, 392–399. doi:10.1016/j.robot.2013.09.012

952 Hall, L., Woods, S., Aylett, R., and Paiva, A. (2006). Using theory of mind methods to investigate
953   empathic engagement with synthetic characters. International Journal of Humanoid Robotics 3, 351–370.
954   doi:10.1142/S0219843606000734

955 Hiah, L., Sidorenkova, T., Romero, L. P., Teh, Y.-F., van Varik, F., Terken, J., et al. (2013). Engaging
956   children in cars through a robot companion. In Proceedings of the 12th International Conference on
957   Interaction Design and Children (ACM), 384–387. doi:10.1145/2485760.2485815

958 Holroyd, A., Rich, C., Sidner, C. L., and Ponsler, B. (2011). Generating connection events for human-
959   robot collaboration. In Proceedings of the 20th IEEE International Symposium on Robot and Human
960   Interactive Communication (RO-MAN) (IEEE), 241–246. doi:10.1109/ROMAN.2011.6005245

961 Hsu, S.-H., Chou, C.-Y., Chen, F.-C., Wang, Y.-K., and Chan, T.-W. (2007). An investigation of
962   the differences between robot and virtual learning companions' influences on students' engagement.
963   In The first IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning
964   (DIGITEL'07) (IEEE), 41–48. doi:10.1109/DIGITEL.2007.10

965 Iacobelli, F. and Cassell, J. (2007). Ethnic identity and engagement in embodied conversational agents. In
966   The international Conference on Intelligent Virtual Agents, eds. C. Pelachaud, J.-C. Martin, E. André,
967   G. Chollet, K. Karpouzis, and D. Pelé. Springer (Springer Berlin Heidelberg), 57–63

968 Inoue, K., Lala, D., Nakamura, S., Takanashi, K., and Kawahara, T. (2016). Annotation and
969   analysis of listener's engagement based on multi-modal behaviors. In Proceedings of the Workshop
970   on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction (ACM), 25–32.
971   doi:10.1145/3011263.3011271

972 Inoue, K., Lala, D., Takanashi, K., and Kawahara, T. (2018). Engagement recognition by a latent character
973   model based on multimodal listener behaviors in spoken dialogue. APSIPA Transactions on Signal and
974   Information Processing 7, e9. doi:10.1017/ATSIP.2018.11

975 Ishii, R. and Nakano, Y. (2010). An empirical study of eye-gaze behaviors: Towards the estimation of
976   conversational engagement in human-agent communication. In Proceedings of the 2010 workshop on
977   Eye gaze in intelligent human machine interaction (ACM), 33–40. doi:10.1145/2002333.2002339

978   Ishii, R. and Nakano, Y. I. (2008). Estimating user's conversational engagement based on gaze behaviors. In
979   <u>Proceedings of the international Conference on Intelligent Virtual Agents</u> (Springer Berlin Heidelberg),
980   200–207

981   Ishii, R., Nakano, Y. I., and Nishida, T. (2013a). Gaze awareness in conversational agents: Estimating a
982   user's conversational engagement from eye gaze. <u>ACM Transactions on Interactive Intelligent Systems</u>
983   <u>(TiiS)</u> 3, 11. doi:10.1145/2499474.2499480

984   Ishii, R., Ooko, R., Nakano, Y. I., and Nishida, T. (2013b). Effectiveness of gaze-based engagement
985   estimation in conversational agents. In <u>Eye Gaze in Intelligent User Interfaces</u> (Springer London).
986   85–110. doi:10.1007/978-1-4471-4784-8_6

987   Ishii, R., Shinohara, Y., Nakano, Y., and Nishida, T. (2011). Combining multiple types of eye-gaze
988   information to predict user's conversational engagement. In <u>The 2nd workshop on eye gaze on intelligent</u>
989   <u>human machine interaction</u>

990   Ivaldi, S., Anzalone, S. M., Rousseau, W., Sigaud, O., and Chetouani, M. (2014). Robot initiative in a
991   team learning task increases the rhythm of interaction but not the perceived engagement. <u>Frontiers in</u>
992   <u>Neurorobotics</u> 8, 5. doi:10.3389/fnbot.2014.00005

993   Ivaldi, S., Lefort, S., Peters, J., Chetouani, M., Provasi, J., and Zibetti, E. (2017). Towards engagement
994   models that consider individual factors in hri: On the relation of extroversion and negative attitude
995   towards robots to gaze and speech during a human–robot assembly task. <u>International Journal of Social</u>
996   <u>Robotics</u> 9, 63–86. doi:10.1007/s12369-016-0357-8

997   Kawamura, K., Rogers, T. E., and Ao, X. (2002). Development of a cognitive model of humans in
998   a multi-agent framework for human-robot interaction. In <u>Proceedings of the first International Joint</u>
999   <u>Conference on Autonomous Agents and Multi-Agent Systems: part 3</u> (ACM), 1379–1386. doi:10.1145/
1000   545056.545135

1001   Keizer, S., Foster, M., Wang, Z., and O., L. (2014). Machine learning for social multi-party human-robot
1002   interaction. <u>ACM Transactions on Interactive Intelligent Systems (TiiS)</u> 4, 14:1–14:32. doi:10.1145/
1003   2600021

1004   Khamassi, M., Velentzas, G., Tsitsimis, T., and Tzafestas, C. (2018). Robot fast adaptation to changes in
1005   human engagement during simulated dynamic social interaction with active exploration in parameterized
1006   reinforcement learning. <u>IEEE Transactions on Cognitive and Developmental Systems</u> 10, 881–893.
1007   doi:10.1109/TCDS.2018.2843122

1008   Kidd, C. D. and Breazeal, C. (2004). Effect of a robot on user perceptions. In <u>Proceedings of the</u>
1009   <u>2004 IEEE/RSJ International Conference on Intelligent Robots and Systems</u> (IEEE), vol. 4, 3559–3564.
1010   doi:10.1109/IROS.2004.1389967

1011   Kim, J., Truong, K. P., Charisi, V., Zaga, C., Evers, V., and Chetouani, M. (2016a). Multimodal detection
1012   of engagement in groups of children using rank learning. In <u>The International Workshop on Human</u>
1013   <u>Behavior Understanding</u> (Springer International Publishing), 35–48

1014   Kim, J., Truong, K. P., and Evers, V. (2016b). Automatic detection of children's engagement using
1015   non-verbal features and ordinal learning. In <u>Workshop on Child Computer Interaction</u>. 29–34

1016   Kim, J., Truong, K. P., and Evers, V. (2018). Automatic temporal ranking of children's engagement levels
1017   using multi-modal cues. <u>Computer Speech & Language</u> 50, 16–39. doi:10.1016/j.csl.2017.12.005

1018   Klotz, D., Wienke, J., Peltason, J., Wrede, B., Wrede, S., Khalidov, V., et al. (2011). Engagement-based
1019   multi-party dialog with a humanoid robot. In <u>Proceedings of the SIGDIAL 2011 Conference</u> (ACL),
1020   341–343

1021   Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A., and Kuzuoka, H. (2007). Museum
1022   guide robot based on sociological interaction analysis. In <u>Proceedings of the SIGCHI Conference on</u>

Human Factors in Computing Systems (ACM), 1191–1194. doi:10.1145/1240624.1240804

Lala, D., Inoue, K., Milhorat, P., and Kawahara, T. (2017). Detection of social signals for recognizing engagement in human-robot interaction. CoRR abs/1709.10257

Leclère, C., Avril, M., Viaux-Savelon, S., Bodeau, N., Achard, C., Missonnier, S., et al. (2016). Interaction and behaviour imaging: A novel method to measure mother–infant interaction using video 3d reconstruction. Translational Psychiatry 6, e816–e816. doi:10.1038/tp.2016.82

Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2014). Empathic robots for long-term interaction. International Journal of Social Robotics 6, 329–341. doi:10.1007/s12369-014-0227-1

Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: a survey. International Journal of Social Robotics 5, 291–308. doi:10.1007/s12369-013-0178-y

Leite, I., McCoy, M., Lohani, M., Salomons, N., Stokes, C., Rivers, S., et al. (2016). Autonomous disengagement classification and repair in multiparty child-robot interaction. In Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (IEEE). doi:10.1109/ROMAN.2016.7745168

Leyzberg, D., Spaulding, S., and Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (ACM), 423–430. doi:10.1145/2559636.2559671

Li, D., Rau, P. P., and Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. International Journal of Social Robotics 2, 175–186. doi:10.1007/s12369-010-0056-9

Li, G., Hung, H., Whiteson, S., and Knox, W. B. (2013). Using informative behavior to increase engagement in the tamer framework. In Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems (International Foundation for Autonomous Agents and Multiagent Systems), 909–916

Malik, N. A., Yussof, H., Hanapiah, F. A., and Anne, S. J. (2014). Human robot interaction (hri) between a humanoid robot and children with cerebral palsy: Experimental framework and measure of engagement. In Proceedings of the 2014 IEEE Conference on Biomedical Engineering and Sciences (IEEE), 430–435. doi:10.1109/IECBES.2014.7047536

Mercado, D. J., Bailly, G., and Pelachaud, C. (2016). Hold my hand, baby: Understanding engagement through the illusion of touch between human and agent. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM (ACM), 1438–1444. doi:10.1145/2851581.2892463

Michalowski, M. P., Sabanovic, S., and Simmons, R. (2006). A spatial model of engagement for a social robot. In The 9th IEEE International Workshop on Advanced Motion Control (IEEE), 762–767. doi:10.1109/AMC.2006.1631755

Moshkina, L., Trickett, S., and Trafton, J. G. (2014). Social engagement in public places: a tale of one robot. In Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (ACM), 382–389. doi:10.1145/2559636.2559678

Najar, A., Sigaud, O., and Chetouani, M. (2016). Training a robot with evaluative feedback and unlabeled guidance signals. In 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 261–266. doi:10.1109/ROMAN.2016.7745140

Nakano, Y. I. and Ishii, R. (2010). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In Proceedings of the 15th international conference on Intelligent User Interfaces (ACM), 139–148. doi:10.1145/1719970.1719990

Nakano, Y. I. and Nishida, T. (2005). Awareness of perceived world and conversational engagement by conversational agents. Informatics for Supporting Social Intelligence , 128

Obaid, M., Gao, Y., Barendregt, W., and Castellano, G. (2017). Exploring users' reactions towards tangible implicit probes for measuring human-robot engagement. In International Conference on Social Robotics (Springer International Publishing), 402–412. doi:10.1007/978-3-319-70022-9_40

Oertel, C., Cummins, F., Edlund, J., Wagner, P., and Campbell, N. (2013). D64: A corpus of richly recorded conversational interaction. Journal on Multimodal User Interfaces 7, 19–28

Oertel, C., Funes Mora, K. A., Sheikhi, S., Odobez, J.-M., and Gustafson, J. (2014). Who will get the grant?: A multimodal corpus for the analysis of conversational behaviours in group interviews. In Proceedings of the 2014 workshop on Understanding and Modeling Multiparty, Multimodal Interactions (ACM), 27–32

Oertel, C. and Salvi, G. (2013). A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. In Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ACM), 99–106

Oertel, C., Scherer, S., and Campbell, N. (2011). On the use of multimodal cues for the prediction of degrees of involvement in spontaneous conversation. In Twelfth Annual Conference of the International Speech Communication Association

Papadopoulos, F., Küster, D., Corrigan, L. J., Kappas, A., and Castellano, G. (2016). Do relative positions and proxemics affect the engagement in a human-robot collaborative scenario? Interaction Studies 17, 321–347

Pitsch, K., Kuzuoka, H., Suzuki, Y., Sussenbach, L., Luff, P., and Heath, C. (2009). The first five seconds: Contingent stepwise entry into an interaction as a means to secure sustained engagement in hri. In The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009. (IEEE), 985–991. doi:10.1109/ROMAN.2009.5326167

Poggi, I. (2007). Mind, hands, face and body: A goal and belief view of multimodal communication (Weidler)

Ponsler, B. (2011). Recognizing engagement behaviors in human-robot interaction

Powers, A., Kiesler, S., Fussell, S., Fussell, S., and Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (ACM), 145–152

Procter, M., Lin, F., and Heller, R. (2016). Improving conversation engagement through data-driven agent behavior modification. In Canadian Conference on Artificial Intelligence (Springer), 270–275

Qureshi, A., Peters, C., and Apperly, I. (2013). Interaction and engagement between an agent and participant in an on-line communication paradigm as mediated by gaze direction. In Proceedings of the 2013 Inputs-Outputs Conference: An Interdisciplinary Conference on Engagement in HCI and Performance (ACM), 8

Ramírez, O. A. I., Khambhaita, H., Chatila, R., Chetouani, M., and Alami, R. (2016). Robots learning how and where to approach people. In 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 347–353

Rani, P. and Sarkar, N. (2005). Operator engagement detection and robot behavior adaptation in human-robot interaction. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation (IEEE), 2051–2056

Rehm, M. and Jensen, M. L. (2015). Accessing cultural artifacts through digital companions: The effects on children's engagement. In Proceedings of the 2015 International Conference on Culture and Computing (Culture Computing) (IEEE), 72–79. doi:10.1109/Culture.and.Computing.2015.44

Rich, C., Ponsler, B., Holroyd, A., and Sidner, C. L. (2010). Recognizing engagement in human-robot interaction. In Proceedings of the 2010 5th ACM/IEEE International Conference on Human-Robot

1112   Interaction (IEEE Press), 375–382

1113 Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., and Olsson, L. (2005). Sustaining
1114   interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an
1115   exploratory study. In IEEE International Workshop on Robot and Human Interactive Communication
1116   (IEEE), 716–722

1117 Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., et al. (2011). Child-robot interaction in
1118   the wild: advice to the aspiring experimenter. In Proceedings of the 13th International Conference on
1119   Multimodal Interfaces (ACM), 335–342

1120 Rudovic, O., Park, H.-W., Busche, J., Schuller, B., Breazeal, C., and Picard, R. W. (2019a). Personalized
1121   estimation of engagement from videos using active learning with deep reinforcement learning. In
1122   Proceedings of IEEE CVPR -AMFG W (IEEE)

1123 Rudovic, O., Zhang, M., Schuller, B., and Picard, R. (2019b). Multi-modal active learning from human data:
1124   A deep reinforcement learning approach. In 2019 International Conference on Multimodal Interaction
1125   (New York, NY, USA: ACM), ICMI '19, 6–15

1126 Šabanović, S., Bennett, C. C., Chang, W.-L., and Huber, L. (2013). Paro robot affects diverse interaction
1127   modalities in group sensory therapy for older adults with dementia. In IEEE 13th International
1128   Conference on Rehabilitation Robotics (ICORR) (IEEE), 1–6

1129 Salam, H., Celiktutan, O., Hupont, I., Gunes, H., and Chetouani, M. (2017). Fully automatic analysis of
1130   engagement and its relationship to personality in human-robot interactions. IEEE Access 5, 705–721

1131 Salam, H. and Chetouani, M. (2015). Engagement detection based on mutli-party cues for human robot
1132   interaction. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)
1133   (IEEE), 341–347

1134 Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis
1135   of affective postures and body motion to detect engagement with a game companion. In Proceedings of
1136   the 6th International Conference on Human-Robot Interaction (ACM), 305–312

1137 Scheeff, M., Pinto, J., Rahardja, K., Snibbe, S., and Tow, R. (2002). Experiences with sparky, a social
1138   robot. In Socially Intelligent Agents (Springer). 173–180

1139 Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). No fair!! an interaction with a cheating robot. In 5th
1140   ACM/IEEE International Conference o Human-Robot Interaction (HRI) (IEEE), 219–226

1141 Sidner, C. L. (2016). Engagement, emotions, and relationships: On building intelligent agents. In Emotions,
1142   Technology, Design, and Learning (Elsevier). 273–294

1143 Sidner, C. L. and Dzikovska, M. (2002). Human-robot interaction: Engagement between humans and
1144   robots for hosting activities. In Proceedings of the 4th IEEE International Conference on Multimodal
1145   Interfaces (IEEE Computer Society), 123

1146 Sidner, C. L. and Dzikovska, M. (2005). A first experiment in engagement for human-robot interaction in
1147   hosting activities. In Advances in Natural Multimodal Dialogue Systems (Springer). 55–76

1148 Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. (2004). Where to look: a study of human-robot engagement.
1149   In Proceedings of the 9th International Conference on Intelligent User Interfaces (ACM), 78–84

1150 Sidner, C. L. and Lee, C. (2003). An architecture for engagement in collaborative conversations between a
1151   robot and humans. Mitsubishi Electric Research Labs TR2003 13

1152 Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans
1153   and robots. Artificial Intelligence 166, 140–164

1154 Sidner, C. L., Lee, C., and Lesh, N. (2003). Engagement when looking: behaviors for robots when
1155   collaborating with people. In Diabruck: Proceedings of the 7th workshop on the Semantic and
1156   Pragmatics of Dialogue (University of Saarland), 123–130

1157 Sidner, C. L., Lee, C., Morency, L.-P., and Forlines, C. (2006). The effect of head-nod recognition in human-
1158   robot conversation. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot
1159   Interaction (ACM), 290–296
1160 Siegel, M., Breazeal, C., and Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on
1161   human behavior. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and
1162   Systems (IEEE), 2563–2568
1163 Sohail, U., Gordon, C., Artstein, R., and Traum, D. (2019). Character initiative in dialogue increases user
1164   engagement and rapport. SemDial Workshop on the Semantics and Pragmatics of Dialogue
1165 Subramainan, L., Mahmoud, M. A., Ahmad, M. S., and Yusoff, M. Z. M. (2016a). A conceptual emotion-
1166   based model to improve students engagement in a classroom using agent-based social simulation.
1167   In Proceedings of the 2016 4th International Conference on User Science and Engineering (IEEE),
1168   149–154
1169 Subramainan, L., Mahmoud, M. A., Ahmad, M. S., and Yusoff, M. Z. M. (2016b). An emotion-based
1170   model for improving students' engagement using agent-based social simulator. International Journal on
1171   Advanced Science, Engineering and Information Technology 6, 952–958
1172 Sun, M., Zhao, Z., and Ma, X. (2017). Sensing and handling engagement dynamics in human-robot
1173   interaction involving peripheral computing devices. In Proceedings of the 2017 CHI Conference on
1174   Human Factors in Computing Systems (ACM), 556–567
1175 Szafir, D. and Mutlu, B. (2012). Pay attention!: designing adaptive agents that monitor and improve user
1176   engagement. In Proceedings of the SIGCHI c+Conference on Human Factors in Computing Systems
1177   (ACM), 11–20
1178 Trinh, H., Shamekhi, A., Kimani, E., and Bickmore, T. W. (2018). Predicting user engagement in
1179   longitudinal interventions with virtual agents. In Proceedings of the 18th International Conference on
1180   Intelligent Virtual Agents (New York, NY, USA: Association for Computing Machinery), IVA '18, 9–16.
1181   doi:10.1145/3267851.3267909
1182 Vaufreydaz, D., Johal, W., and Combe, C. (2016). Starting engagement detection towards a companion
1183   robot using multimodal features. Robotics and Autonomous Systems 75, 4–16
1184 Vázquez, M., Steinfeld, A., Hudson, S. E., and Forlizzi, J. (2014). Spatial and other social engagement cues
1185   in a child-robot interaction: Effects of a sidekick. In Proceedings of the 2014 ACM/IEEE International
1186   Conference on Human-Robot Interaction (ACM), 391–398
1187 Xu, Q., Li, L., and Wang, G. (2013). Designing engagement-aware agents for multiparty conversations. In
1188   Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (ACM), 2233–2242
1189 Youssef, A. B., Varni, G., Essid, S., and Clavel, C. (2019). On-the-fly detection of user engagement
1190   decrease in spontaneous human-robot interaction. International Journal on Social Robotics
1191 Zaga, C., Lohse, M., Truong, K. P., and Evers, V. (2015). The effect of a robot's social character on
1192   children's task engagement: Peer versus tutor. In Proceedings of the International Conference on Social
1193   Robotics (Springer International Publishing), 704–713