Chapter 7
Fostering User Engagement in Face-to-Face Human-Agent Interactions: A Survey

**Chloé Clavel, Angelo Cafaro, Sabrina Campano
and Catherine Pelachaud**

Abstract Embodied conversational agents are capable of carrying a face-to-face interaction with users. Their use is substantially increasing in numerous applications ranging from tutoring systems to ambient assisted living. In such applications, one of the main challenges is to keep the user engaged in the interaction with the agent. The present chapter provides an overview of the scientific issues underlying the engagement paradigm, including a review on methodologies for assessing user engagement in human-agent interaction. It presents three studies that have been conducted within the Greta/VIB platforms. These studies aimed at designing engaging agents using different interaction strategies (alignment and dynamical coupling) and the expression of interpersonal attitudes in multi-party interactions.

7.1 Introduction

One of the key challenges of human-agent interaction is to maintain user engagement. The design of engaging agents is paramount, whether in short-term human-agent interactions, or for building long-term relations between the user and the agent.

C. Clavel (✉) · A. Cafaro · C. Pelachaud
Telecom-ParisTech, LTCI, CNRS, Université Paris-Saclay, 46 rue Barrault, 75013 Paris, France
e-mail: chloe.clavel@telecom-paristech.fr

A. Cafaro
e-mail: angelo.cafaro@telecom-paristech.fr

C. Pelachaud
e-mail: catherine.pelachaud@telecom-paristech.fr
URL:http://www.tsi.telecom-paristech.fr/mm/en/themes-2/greta-team

S. Campano
Lab Object'ive, Object'ive, 96 98 rue de Montreuil, 75011 Paris, France
e-mail: scampano@object-ive.com

Many applications of human-agent interaction such as tutoring systems [45], ambient assisting living [12, 61] or virtual museum agents [24, 31, 58, 72] show the importance of the engagement paradigm. In ambient assisted living and tutoring systems, for example, the challenge is to maintain user engagement over many interactions, while in museum applications, the key issue is to invite visitors to interact and keep them engaged during the interaction for as long as possible.

In the human computer interaction literature, the issue of engagement is addressed from different angles. An interesting way of structuring such a literature is to rely on the distinction provided by Peters and colleagues [105], who distinguished the two following components underlying the engagement process: the *attentional involvement* and the *emotional involvement*. It is important to notice that, even though some studies focus more on one of these two components, they interleave, as the attention is driven by emotions. The definitions provided by Sidner and Dzikovska [125]—"*the process by which individuals in an interaction start, maintain and end their perceived connection to one another*"—focus on the *attentional involvement*, while [47, 63] focus on emotional engagement. In particular, in [63] they concentrate on empathic engagement—"*Empathic engagement is the fostering of emotional involvement intending to create a coherent cognitive and emotional experience which results in empathic relations between a user and a synthetic character*". Another major distinction provided by Bickmore et al. [14] differentiates short-term versus long-term engagement. The former deals with user engagement in performing a task[1] while interacting with the agent. The latter implies much longer periods of interactions with the system and concerns the degree of involvement of the user over time.

The present chapter provides a review of the various literature dealing with the common objective of fostering user engagement in human-agent interactions (Sect. 7.2). The literature calls in different research fields ranging from social signal processing and affective computing to dialogue management and perceptive studies. Then, we focus on the description of examples of studies carried out inside a common platform—the Greta platform—(Sect. 7.3). Finally, we conclude and provide some tracks for the future design of engaging agents.

7.2 Designing Engaging Agents—State of the Art

"*Embodied Conversational Agents (ECAs) are virtual anthropomorphic characters which are able to engage a user in real-time, multimodal dialogue, using speech, gesture, gaze, posture, intonation and other verbal and nonverbal channels to emulate the experience of human face-to-face interaction*" [32]. Following this definition, in this section, we review the different research themes and issues involved in the usage of ECAs in order to foster user's engagement. These issues are represented in

---

[1]In this case, the task can just be interacting with the agent.
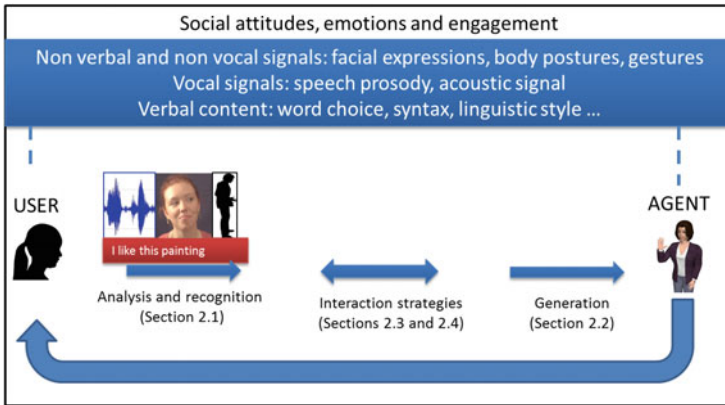
Fig. 7.1 The diagram shows the main research themes in the area of designing engaging agents. The focus is on the generation (*agent*) and the recognition (*user*) of socio-emotional behavior

the diagram in Fig. 7.1 which also shows the main multimodal channels adopted in human-agent interaction: non verbal and non vocal signals such as facial expressions, vocal signals such as prosody, and verbal content, such as word choice.

First, user's verbal and non verbal behavior should be taken into account by analysing the different modalities, on the one hand (Sect. 7.2.1) and the ECA should express relevant socio-emotional behavior through these modalities, on the other hand (Sect. 7.2.2). Then, multimodal dialogue management should be considered. In particular, we should address how to implement socio-emotional interactions between the user and the agent (Sects. 7.2.3 and 7.2.4). Finally, evaluation issues are paramount in human-agent interactions. The main questions are how to measure the impact of the design of engaging agent on user's impression (Sect. 7.2.5) and how to evaluate the 'success' of the interaction by analyzing the user engagement (Sect. 7.2.6).

7.2.1 Taking into account Socio-Emotional Behavior

ECAs aim at facilitating the socio-emotional interaction between the human and the machine. The socio-emotional aspect is a main prerequisite for a fluent interaction and thus for user's engagement. It relies on the development of agents endowed with socio-emotional abilities i.e. agents that are able to take into account user's social attitudes and emotions.

The user's expression of socio-emotional behavior can be both verbal and non-verbal (acoustic component of speech, facial expressions, gesture, body posture). Existing studies focused on the acoustic features such as prosody, voice quality or spectral features [27, 38, 124] and more generally on non-verbal features (posture,

gesture, gaze, or facial expressions [89]) for emotion detection. Recently, the analysis of emotion has been integrated in a more general domain which considers also social behavior: the domain of social signal processing [138].

The analysis of non-verbal emotional content is widespread in human-agent inter-action [96, 123, 128, 143]. The detection of the user's socio-emotional behaviors takes part in the inputs of agent's socio-emotional interaction strategies. It can also be considered as an input of a global user model for the building of long-term relations between the user and the agent [14]. It is strongly linked to user engagement: the detection of negative emotional states of the user is considered as a premise of user disengagement in the interaction [121]. Besides, avoiding (and thus, detecting) user frustration is also a key challenge to improve user learning in tutoring systems [77]. This statement is reinforced in [102] where they claim that it is important to consider students' achievement goals and emotions in order to promote their engagement and in [45], where they provide an *Affective auto-tutor agent* able to detect student's boredom and engagement. It is also interesting to note that some studies are not con-sidering the socio-emotional level to detect engagement or disengagement but prefer to consider directly the signal level such as face location [17]. This last type of studies is efficient to detect user disengagement such as quitting the interaction. However, analyzing socio-emotional behavior as a cue of user engagement or disengagement supports the detection of subtler changes in the user engagement process.

The verbal content of emotions corresponds more to sentiment, opinion (see [87] for a discussion about the different terminologies) and attitude [120]. It begins to be integrated for the analysis of user socio-emotional behaviors. Natural language processing is yet not necessarily restricted to the analysis of the topic of user's utterance and can now give access to these socio-emotional cues [37]. In this way, in [142] they provided a system based on verbal cues that distinguished neutral, polite and frustrated user states. In [128], they proposed a classification between positive *vs.* negative user sentiment as an input of human-agent interaction. In [74] they provided a model of user's *attitudes* in verbal content grounded on the model described in [83] and dealing with the interaction context: as the previous agent's utterance can trigger or constrain an the user's expression of attitude, its target or its polarity, the model of the semantic and pragmatic features of agent's utterance is used to help the detection of user's attitude. Relying on the joint analysis of the agent's and the user's adjacent utterances, in [75] they provide a system able to detect user's like and dislike and devoted to the improvement of the social relationships between the agent and the user.

7.2.2 Generation of Agent's Socio-Emotional Behavior

An engaging agent in addition to perceiving the user's level of engagement should also be capable of maintaining it by exhibiting the appropriate socio-emotional behav-ior during the interaction. Two major issues arise regarding (1) the type of behavior

to display and (2) when it should be exhibited. In this section we start by discussing the first issue continuing with the second one in Sect. 7.2.3. In particular, we describe several approaches adopted for the generation of multimodal behavior supporting the expression of social attitudes and emotions.

Expression of social attitudes As for modeling multimodal behaviors associated to social attitudes, several approaches rely on the Interpersonal Circumplex of attitudes proposed by Argyle [5] and on the correlation between specific behavior patterns and the expression of attitudes according to Burgoon and colleagues [22]. Two dimensions are considered, namely liking (or affiliation) and dominance (or status) [24, 36, 56, 76, 112]. Multimodal behaviors including gaze and head movement, body orientation, facial expression, use of personal space can be exhibited for expressing different attitudes. In order to find out and to be able to model such a correlation between behaviors and attitudes, different methodologies have been proposed. Bickmore and colleagues [16] have incorporated findings from social psychology to specify the behavior of their relational agent Laura. Experimental studies have been designed where human subjects were asked to indicate their perception of social attitudes of virtual agents [7, 8]. They show the importance of flirting tactics through gaze behaviors and expressive mimics in order to establish a first contact between the agent and the user [7] and that gaze behaviors and the linguistic expression of disagreeableness have a significant effect on the perception of dominance [8]. Several researchers collected and analyzed corpora of interacting human participants [36, 76], allowing the extraction of behaviors patterns linked to social attitudes. Ravenet et al. [112] used a crowd sourcing method asking human subjects to design agents with different attitudes by selecting multimodal behaviors through an interactive interface. The perception of a behavior (e.g. a smile) may vary depending on the chosen timing and context in which it is exhibited. For example, a smile followed by a gaze shift conveys a different attitude compared to a smile followed by a leaning toward the interlocutor. According to this, Chollet and colleagues [36] proposed to model the expression of an attitude as a sequence of behaviors. From the analysis of a corpus which has been annotated on two levels, attitudes and multimodal behaviors, sequences of multimodal behaviors linked to an attitude, are extracted using a sequence mining approach as suggested in [129].

While several studies focused on expressing social attitudes in face-to-face interaction, others have looked at the expression of an attitude by an agent in a multi-party interaction (e.g. in a conversing group) [55, 76, 112]. In such scenarios, the behavior exhibited by an agent needs to take into account the behaviors exhibited by other participants (for example in a group conversation) in order to compute the intended attitudes. In the project Demeanour [56], they modeled the posture and the gaze direction of virtual agents interacting with each other. The agents adapted their gaze direction, in particular their amount of mutual gaze, based on their attitudes toward each other. Ravenet et al. [114] went further in this direction and proposed a model of turn-taking. Depending on the attitudes toward each other (that may or may not be symmetrical), the agents adapt their spatial relations, their body orientation, gesture quality. They also change their manner to take the speaking turn or to handle inter-

ruption. For example, an agent that is dominant towards another has the tendency to interrupt the latter while it holds the speaking turn.

Expression of emotions Early ECAs models (c.f. [11, 103]) focused on the six prototypical expressions of emotions namely: anger, disgust, fear, joy, sadness and surprise (see [49]). However these expressions are barely used in interaction. Later on, researchers focused on endowing an agent with more subtle and more varied expressions of emotions. The proposed models can be distinguished by the theoretical model of emotions they used. Three main approaches can be reported: discrete emotion theory, dimension theory and appraisal theory. Computational models of the expressions of emotions rely on one of these models.

The discrete emotion theory introduced by Ekman [48] and Izard [68] claims that there is a small set of primary emotions that are universally produced and recognized. The expressions of these emotions can be blended as suggested by Ekman and Friesen [49]. As above mentioned, early models were built using findings from these theoretical models. Models of expression blending have been proposed in [21, 91]. Fuzzy logic was used to blend expressions of two emotions on the different parts of the face.

The dimensional theory describes emotions along a continuum over two [107, 117], or three [84], or even four dimensions [51]. The most common dimensions are pleasure, arousal and dominance. Emotions are no longer referred by a label (e.g., relief, regret) but by their coordinates in space. Computational models relying on dimensional representations make use of this continuum. They propose to create new expressions as the blending of facial expressions of known emotions placed in the 2D or 3D space. One of the first models was proposed by Ruttkay et al. [118]. The authors have developed a tool called Emotion Disc. Expressions of the six prototypical emotions are placed around a circle. The distance from the center to the outer of the circle indicates the intensity of the expression. A new expression can be created as a linear interpolation of these prototypical expressions. Other researchers [3, 136] proposed to calculate a new expression by interpolation between the closest known expressions. The interpolation can be done in 2D [136] and in 3D [3] space. Such approaches allow computing intermediate expressions from existing ones.

In [18, 59] they followed a very different approach. While the methods just presented are based on the prototypical expressions, these latter authors create a large set of facial expressions by composing randomly actions units defined with FACS [50]. Then they ask participants to rate the expressions along 2D [59] or 3D [18] space.

Appraisal theory views emotions as arising from the evaluation of an event, object, or person along different dimensions. In particular, the Componential Process Model (CPM) introduced by Courgeon et al. [119] makes predictions between how an event is appraised and the facial response. Few attempts [39, 97], have been made to implement how the facial responses are temporally organized to create the expression of emotion. Thus the expression of emotion does not correspond to a full-blown expression that arises in a block; rather it is made of a sequence of signals that arises and is composed on the face. In [92, 141], they pushed forward this idea. Based on a corpus analysis where multimodal behaviors have been annotated, the

authors extract sequences of behaviors linked to emotions either manually [92] or automatically [141] using the T-pattern model developed by [81]. From the extraction of the data, Niewiadomski and colleagues [92] defined a set of rules that encompasses the spatial and temporal constraints of the signals in the sequences. Such models allow generating the expressions of emotions as sequences of temporally-ordered multimodal signals.

7.2.3 Socio-Emotional Interaction Strategies

In addition to the requirement of taking into account user socio-emotional behavior, on the one hand and to generate believable and engaging socio-emotional behavior of the agent, on the other hand, HCI requires to define the socio-emotional strategies linking the user input to the agent output. Existing strategies do not always have the explicit goal of fostering user engagement. In this paragraph, we focus on examples of strategies that have been explicitly used to foster user's engagement or to improve feelings of *rapport*, a concept which is strongly linked to engagement [14].

Providing backchannels and feedbacks is a key strategy for maintaining user engagement by providing agent's listening behaviors [73]. Thus, in the study of D'Mello and Graesser [45], the Auto-tutor agent provided feedback in order to help students to regulate their disengagement (boredom, etc.). In [122], the agent was able to generate multimodal backchannel (smile, nod, and verbal content) when it is listening to the user, and the timing of the backchannel—that is when to trigger the backchannel—was provided by probabilistic rules. In [135], they provided another rule-based model in order to predict when a backchannel has to be triggered as a reaction to prosody and pause behaviors. In [86] they used sequential probabilistic models, an interesting method to predict jointly when and how to generate backchannels in the listening phase of the agent. The timing issue of backchannel is close to the issue tackled in turn-taking strategies, that is, when the agent has to take or give the floor. As described in Sect. 7.2.5, researchers presented different turn-taking strategies and evaluated their role on the user's impressions [79, 80].

Politeness strategies are also associated with the concept of engagement. They provide to the agent a social intelligence [139] and allow it to be perceived as more engaged in the interaction [57]. In [4], politeness strategies were used as an answer to the expression of negative emotional states by the user to adjust the politeness level of their virtual guide. The more the interlocutor is in a negative emotional state, the more the guide has to be polite. However, Campano et al. [28] showed that in certain situations such as in video games, the agent has to express impoliteness to be more believable.

Endowing agents with humor may be a smart answer when the user is confused in front of some dysfunctions of the interaction system. Dybala and colleagues [47] proposed a humor-equipped casual conversational system (chatbot) and demonstrated that it enhances the user's positive engagement/involvement in the conversation.

A last example of smart strategies dedicated to improve user engagement is the management of agent's surprise. Bohus and Horvitz [17] proposed to communicate the robot's surprise when the user seemed to be disengaged in the interaction by using linguistic hesitation.

7.2.4 Alignment-Related Processes

Alignment [106] of ECA's behavior on the user is another strategy for improving user's engagement. Various approaches are used to design alignment processes or similar processes. These processes differ on the way they integrate the temporal and dynamic aspects. For example, mimicry is defined as the direct imitation of what the other participant produces [10], while synchrony is defined as the dynamic and reciprocal adaptation of temporal structures of behaviors between interactive partners [42]. The processes also differ on the levels at which they occur. At the lowest level, the processes concern the imitation of different modalities: body postures [34], gestures [85], accent and speech rate [54], phonetic realizations [98], word choice [53], repetitions [10, 140], syntax [19] and linguistic style [90].

The higher levels are mental, emotional or cognitive levels. Emotional resonance [60], affiliation [130]—that is alignment on users opinion or attitude (see Sect. 7.3.3)—and alignment at the level of concepts [20] are examples of high-level processes. But the different levels interleave. For example, copying gestures can be viewed as a way to establish and maintain an empathetic connection [33].

Alignment-related processes have been largely studied through linguistic studies dedicated to the observation of corpora. However, recent years have seen an increase of interest in the implementation of alignment-related processes in human-computer interactions, and in human-agent interactions, in particular. Implementations of alignment strategies in human-computer dialogues concerned mainly alignment on lexical and syntactic choices [23], while the human-agent *face-to-face* interactions further implementations of non verbal alignment. It is also interesting to notice that the terminology used in human-agent interaction is slightly different from the one used in corpus studies. In human-agent interaction, the terminology includes terms such as: mimicry [64], coordination [60, 71], synchrony [42], social/emotional resonance [60, 71], emotional mirroring [1], and dynamical coupling [109].

Some researchers attempted to implement complex alignment-related processes in simulated agent-agent interactions, dealing with social resonance and coverbal coordination in [71] and smile reinforcement between two virtual characters [95].

In summary, the literature on the design of socio-emotional interaction strategies is plentiful in the ECA community. Sophisticated interaction strategies such as alignment-related processes are even more frequent and begin to be effectively integrated in ECA platforms.

7.2.5 Impact on User's Impression

Users' evaluation of agents' behavior and interaction strategies are fundamental for designing believable and engaging agents. Recent studies focused on evaluating agents' nonverbal multimodal behavior during the first interaction and, some in particular, focused on the very initial moments. In light of Bickmore's distinction between short *vs.* long term engagement (cf. Sect. 7.1), evaluating the user's impressions of an agent addresses usability issues in both short and long term interactions. The idea is that a more engaging agent during the first interaction is likely to form a positive impression and be accepted by the user, thus promoting further interactions [9, 24].

There is a great deal of information that can be picked from observing an agent's multimodal behavior during the interaction, some of the relevant studies presented in this section mainly dealt with the user's impressions of the agent's friendliness, dominance, agreeableness, warmth and competence. Therefore, the emphasis has been on agent's characteristics such as interpersonal attitude towards the user, personality and skill level in a selected context (e.g. competence), that can be extrapolated from brief observations of multimodal behavior.

Maat et al. [79, 80] showed how a realization of a simple communicative function (for managing the interaction) could influence users' impressions of an agent. They focused on impressions of personality (agreeableness), emotion and social attitudes (i.e. friendliness) through different turn-taking strategies in human face-to-face conversations applied to their virtual agents in order to create different impressions of them. Fukayama and colleagues [52] proposed and evaluated a gaze movement model that enabled a virtual agent to convey different impression to users. They used an "eyes-only" agent on a black background and the impressions they focused on were affiliation (friendliness, warmth) and status (dominance, assurance). Similarly, Takashima et al. [132] evaluated the effects of different eye blinking rates of virtual agents on the viewers subjective impressions of friendliness (a blink rate of about 18 blink/min for the avatar makes a friendly impression), nervousness (higher blink rates reinforced nervous impressions) and intelligence (lower blink rates gave intelligent impression).

Niewiadomski and colleagues [93] analyzed how the emotional multimodal behavior of a virtual assistant expressing happiness, sadness and fear influenced user's judgments of agent's warmth, competence and believability. In particular, socially appropriate emotions expressed by the agent led to higher perceived believability. Then they also found that the perception of agents' believability was highly correlated to the two major socio-cognitive dimensions of warmth and competence.

In [25, 26] they investigated how users interpreted an agent's nonverbal greeting behavior (i.e. smile, gaze and proxemics) in a first encounter in terms of friendly interpersonal attitudes and extraverted personality [26]. In a follow-up study they discovered that a friendly interpersonal attitude expressed with more smiling and gazing at user is more relevant than expressing extraversion with proxemics behavior when it comes to decide whether to continue the interaction with an agent [25].

Bergmann et al. [9] studied how appearance and nonverbal behavior, in particular gestures, affected the perceived warmth and competence of virtual agents over time. Their goal was to study how warmth and competence ratings changed from a first impression after a few seconds to a second impression after a longer period of human-agent interaction, depending on manipulations of the virtual agent's appearance (robot-like character vs. anthropomorphic virtual agent) and gestural behavior (absent vs. present). Results indicated that impressions of warmth changed over time and depended on the agent's appearance. Evaluations of competence also changed but seemed to depend more on gestural behavior.

Virtual and robotic conversational agents have been deployed in public spaces for field studies. These deployments allowed researchers to move from the controlled laboratory settings to a more natural real life environment. Researchers have examined different engagement strategies in first user-agent encounters in such locations (e.g. museums, reception halls) where a multitude of users is present. Experiments conducted in these settings yielded more natural data, but they face a challenging environment that can be noisy, for example, in a museum there could be distracting or competing stimuli.

Gockley and colleagues [58] built a robot receptionist installed in a hall at Carnegie Mellon University, in USA. Valerie was able to give directions to visitors and look up the weather forecast while also exhibiting a compelling personality and character to encourage multiple visits over extended periods of time. The robot classified users according to an attentional zone based on their proximity and orientation (e.g. "engaged" visitors were close by the exhibit but not facing directly it).

Kopp et al. [72] installed Max in the Heinz Nixdorf Museums Forum (HNF), in Germany. Max was projected on a life-size screen and it was designed for being an enjoyable and cooperative interaction partner. It was able to engage with visitors in natural face-to-face conversations with a German voice accompanied by appropriate nonverbal behaviors such as facial expressions, gaze, and locomotion.

Cafaro et al. [24] conducted a study on Tinker at the Boston Museum of Science. Tinker was a human-sized conversational agent displayed as a cartoonish anthropomorphic robot that was capable of describing exhibits in the museum, giving directions, and discussing technical aspects of its own implementation. It used nonverbal conversational behavior, empathy, social dialogue, reciprocal self-disclosure and other relational behavior to establish social bonds with users. Tinker exhibited different greeting behaviors to approaching visitors (e.g. smiling behavior for friendliness). The visitors' commitment to interact with the agent was taken as behavioral measure of user's engagement. In the specific context of a first approach towards the exhibit, this measure was obtained by counting four possible actions from the moment the visitor was in the exhibit's area to the beginning of the interaction with the agent. These possible actions were: (i) walking past the exhibit, or (ii) finishing the approach towards the exhibit, (iii) following some instructions provided by Tinker on how to interact and (iv) effectively starting a conversation. There weren't significant differences among the groups receiving different greeting styles (i.e. no reaction, friendly and unfriendly), however trends seemed to indicate that the friendly version encouraged visitors to undertake more actions.

In summary, laboratory and field studies have been conducted to evaluate user's impressions of virtual and robotic agents. These studies focused on particular dimensions of first impressions such as interpersonal attitude and personality in order to make agents more engaging and accepted for long-term interactions.

7.2.6 Methodologies for Evaluating User Engagement in Human-Agent Interactions

So far we discussed state-of-the-art techniques and strategies for designing engaging ECAs in face-to-face or multi-party interactions with users. We also briefly reviewed some studies aimed at evaluating the impact of agents' exhibited socio-emotional behavior on user's impressions of ECAs. These studies focused on specific dimensions of user's impressions (e.g. agent's interpersonal attitude, competence, warmth) that are likely to improve the level of engagement with the agent. In this section, we move further from the mere assessment of users' first impressions by providing a brief survey on existing methodologies adopted by researchers to assess the user's engagement.

User engagement with an ECA, in general, can be measured via user self-reports (i.e. *subjective*); by monitoring the user's responses, tracking the user's body postures, intonations, head movements and facial expressions during the interaction (i.e. *objective*); or by manually logging behavioral responses of user experience (i.e. *behavioral* or *annotated*). This reflects a common categorization in experimental design [40]. A researcher could attempt to adopt any of the three above-mentioned approaches (or even combinations of those) to capture engagement.

Prior to providing some examples adopting these different methodologies, we should consider another factor that affects the assessment of engagement. In particular, the time window within which users should be asked (or measured) to express (or being detected) the level of engagement with an ECA. Reporting on paper or on a digital questionnaire sheet is the most popular approach to subjective assessment for user engagement, either asking the user upon showing a stimulus or at the end of a series of stimuli. However, there are two extremes that could be considered. One is focusing on real-time assessments *during* the interaction (mostly suitable, for example, when taking objective physiological measurements). On the other hand, there could be a longitudinal assessment taken over repeated interactions in a time span that might cover days, weeks or months of user-agent interactions. We refer to these different timings for assessing engagement as: *within-interaction* (during the interaction, for example at the end of the agent's turn), *end-interaction* and *over-several-interactions* (i.e. in longitudinal studies over multiple interactions).

Subjective assessments of engagement to date have been obtained through questionnaires with closed or open questions, or with structured interviews. Methods such as self-report closed questionnaires are constraining users into specific questionnaire items yielding data that can be easily used for analysis. However, there could

be experimental noise in the responses. For example, participants might be biased after repeated interactions and there could be users' memory limitations about the perceived agent's behavior if asked at the end of interaction (*post-stimuli*), and self-deception (i.e. user not providing the true responses). Example of questionnaires adopted for measuring engagement can be found in an evaluation study presented in [67] and as a dimension of the Temple Presence Inventory (TPI) [78]. In [126], for instance, they adapted the TPI dimensions for studying user-robot engagement. Furthermore, in [46] they developed the Post-Lecture Engagement Questionnaire that required participants to self-report their engagement levels after each lecture. There were three questions which asked participants to rate their engagement at the beginning, middle, and end of each lecture. Participants indicated their ratings on a six-point scale ranging from (1) very bored to (6) very engaged.

Interviews may offer richer information, but the nature of these less structured data compared to quantitative data is harder to analyze. Example of these assessments are structured interviews or free text responses (leading to, for example, adjective analysis). Traum and colleagues [134] measured visitors' engagement when interacting with a pair of museum agents by adopting a mixed approach with both a self-report questionnaire and interviewing subjects at the end of the interaction.

The administration of subjective assessments is usually done at the end of the interaction or over several interactions. It might be intrusive and hard to obtain within-interaction (i.e. questions appearing during the interaction).

Finally, an example of longitudinal assessment with focus on building a working alliance between user and agent in the health domain can be found in [13].

Objective studies rely on automatic detection of physiological [35], verbal or non verbal signals that can be linked to engagement. They can be driven both within the interaction, at the end of the interaction or longitudinally (over several interactions). Analysis of user engagement within the interaction can be provided by automatic analysis such as the one described in Sect. 7.2.1 by analyzing speech prosody or body postures, emotions and attitudes in order to infer user engagement. Unlike subjective self-reports, automatic analysis provides both information on the evolution of the user engagement *within* the interaction and global evaluation of user engagement. Simple automatic measurements at the interaction level can also be provided: Bickmore and colleagues [15] measured the total time in minutes each visitor spent with a relational agent installed in a museum.

Another way to assess user engagement within the interaction and to capture its evolution along the interaction is to carry out *behavioral studies*. Sidner et al. [126], thus provided annotations of videotaped interactions between the user and a robot including the duration of the interaction, the amount of shared looking (looking the same object), mutual gazes (looking at each other), looking at the robot during the humans turn and overall amount of time the user spent looking at the robot. In [88] they describe a study in which a user interacted with an ECA and an external observer watched this interaction. In addition, a push-button device was given to both the user and the observer. The user was instructed to press the button when the agents

explanation was boring and the user would like to change the topic. The observer was instructed to press the button when the user looked bored and distracted from the conversation.

A strength of behavioral and objective studies is their lack of intrusion into the user-agent interaction experience. However, objective studies can be exposed to detection errors, for example when automatically recognizing user's multimodal behavior and behavioral studies are subjected to labelers subjectivity even though they are more shielded from subject bias than subjective studies.

Like subjective studies, objective and behavioral studies are relevant for a longitudinal assessment of engagement over several interactions, an issue which is especially important for applications such as assisted living [14].

7.2.7   Summary of the Key Points for the Design of Engaging Agent

The socio-emotional component has a key role in the design of engaging agent. The literature on users' emotion recognition and on the generation of the agent's emotional behavior begins to have a quite long tradition and to offer a range of satisfactory tools for non-verbal aspects. Further work needs to be done concerning the analysis of the user's verbal socio-emotional content and the use of user socio-emotional behavior in socio-emotional interaction strategies. Besides, the integration of social component with the generation of agents' social stances is more recent and is a promising contribution to the engagement paradigm in human-agent interaction. Next section provides a summary of studies dealing with these three scientific challenges: integration of the verbal content (Sect. 7.3.3) and of the non verbal content in socio-emotional interaction strategies (Sect. 7.3.2); and the expression of social stances in multiparty group interaction (Sect. 7.3.4).

7.3   Overview of Studies Carried Out in GRETA and VIB

The design of engaging agents has been implemented by several studies around a same platform that makes it possible to integrate the different modules required for an engaging human-agent interaction—from the detection of user socio-emotional behavior to the generation of agent socio-emotional behaviors: the Greta system and VIB platform. In this section, we first present the architecture of the Greta system, then we show the extension of this system in the VIB platform, and finally, we present three different studies dedicated to foster user engagement that have been implemented in VIB/Greta. The first two studies deal with computational models of alignment-related processes (dynamical coupling and alignment) as described in Sect. 7.2.4. In particular, Sect. 7.3.2 shows how dynamical coupling can improve user

experience and contribute to user engagement and Sect. 7.3.3 focuses on alignment strategies and their impact on user engagement. The third study focuses on user experience and engagement in multiparty interactions with conversing group of virtual agents.

7.3.1 Greta System and VIB Platform

The Greta system allows a virtual or physical (e.g. robotic) embodied conversational agent to communicate with a human user [94, 95]. The global architecture of the system is depicted in Fig. 7.2. It is a SAIBA compliant architecture (SAIBA is a common framework for the autonomous generation of multimodal communicative behavior in Embodied conversational agents [70]). The main three components are: (1) an *Intent Planner* that produces the communicative intentions and handles the emotional state of the agent; (2) a *Behavior Planner* that transforms the communicative intentions received in input into multimodal signals and (3) a *Behavior Realizer* that produces the movements and rotations for the joints of the ECA.

A *Behavior Lexicon* (i.e. Agent Behavior Specification in Fig. 7.2) contains pairs of mappings from communicative intentions to multimodal signals. The Behavior Realizer instantiates the multimodal behaviors, it handles the synchronization with speech and generates the animations for the ECA.

The information exchanged by these components is encoded in specific representation languages defined by SAIBA. The representation of communicative intents is done with the Function Markup Language (FML) [65]. FML describes communicative and expressive functions without any reference to physical behavior, representing in essence what the agent's mind decides. It is meant to provide a semantic descrip-
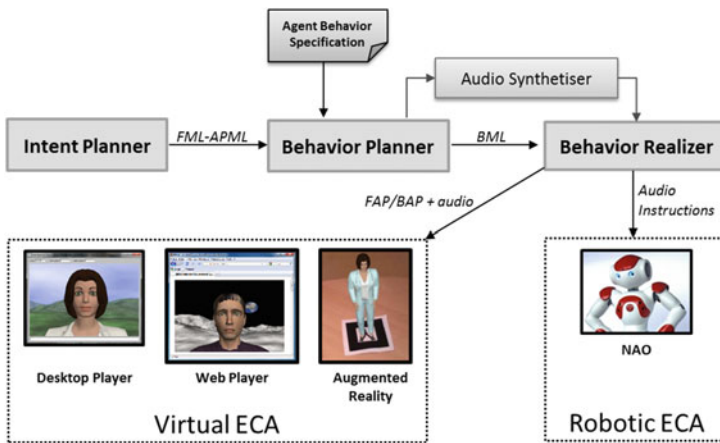

Fig. 7.2 The Greta system

tion that accounts for the aspects that are relevant and influential in the planning of verbal and nonverbal behavior. Greta uses an FML specification named *FML-APML* and based on the Affective Presentation Markup Language (APML) introduced by [41]. FML-APML tags encode the communicative intentions following the taxonomy defined in [108], where a communicative function corresponds to a pair (*meaning*,*signal*). The meaning element is the communicative intent that the ECA aims to accomplish, whereas the signal element indicates the multimodal behavior exhibited in order to achieve the desired communicative intent.

The multimodal behaviors to express a given communicative function to achieve (e.g. facial expressions, gestures and postures) are described by the Behavior Markup Language (BML) [70, 137].

The Greta system has been embedded in the Virtual Interactive Behavior (VIB) platform [99]. An overview of the VIB architecture is shown in Fig. 7.3. VIB enhances Greta with additional components that allow the ECA to detect its environment (i.e. *Perceptive Space* in Fig. 7.3), and to interact with the user while constantly updating the agent's mental and emotional states. Thus, an ECA's mental state includes information such as beliefs, goals, emotions and social attitudes.

The agent's emotional state is computed with the FaTiMa emotion model by [43]. A dialogue manger computes the utterances spoken by the agent as a function of both its mental state and previous verbal content exchanged with the user. Currently VIB integrates the DISCO dialogue manager developed by [116]. The output of this component is sent to the agent's intent planner.
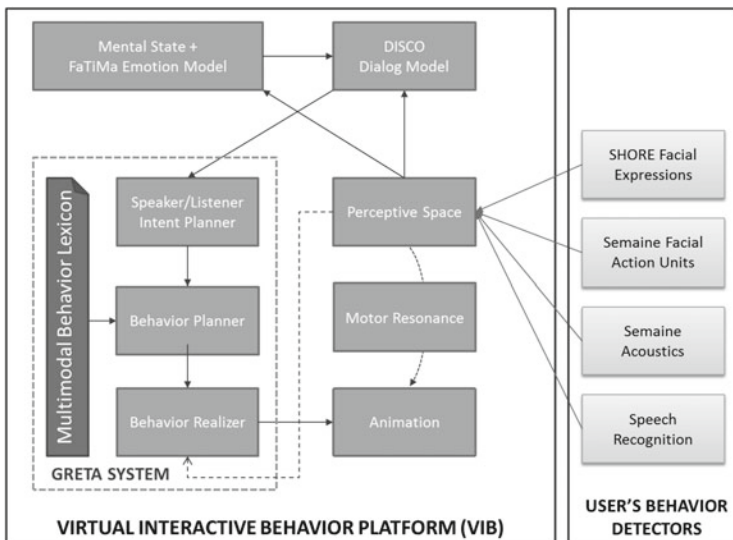

Fig. 7.3 Global architecture of the VIB platform

Different external tools plugged-in the VIB platform (i.e. SHORE facial expressions, SEMAINE facial action units and acoustics, and speech recognition as shown on the right side of Fig. 7.3) allow an agent to detect and interpret user's audio-visual input cues captured with devices such as cameras, Microsoft's Kinect and microphones. These information are provided to the agent via the *Perceptive Space* module. A direct link between this module and the *Behavior Realizer* allows the agent to exhibit reactive behaviors by quickly producing the behavior to exhibit in response to user's behavior, as for back-channels for example.

Finally, the *Motor resonance* manages the direct influence of the socio-emotional behaviors of the user (agent perceptive space) to the ones of the agent (agent production space) without cognitive reasoning. In particular, it allows the ECA to dynamically mimic the behavior of the user.

7.3.2   Modeling Dynamical Coupling

The study presented in this section focused on the *Motor Resonance* module in the GRETA platform. This study is about mirroring of human laughter by an ECA during an interaction. We refer to this process as *dynamical coupling*. The tool supporting the modeling of dynamic coupling in the platform can be used for other communicative functions. An interface allows us to connect the detected users inputs to ECA's animation parameters through a neural network.

Laughter is a social signal that has many functions in dialogue. For example, it allows someone to display a feeling of pleasure consecutively to positive events, such as receiving compliments [110] or perceiving a humorous stimulus. Laughter also serves to hide one's embarrassment [66], or to be cynical. It helps to create social bonds within groups [2], and regulates the speech flow in conversation [110]. These socio-emotional communicative functions are important in interaction. It is then important to enable ECAs to laugh in order to improve the quality of human-agent interaction, and to enhance user involvement. To this end, we defined a model of laughter [44], currently integrated in the GRETA architecture, and we conducted an evaluation that explores the role of laughter mirroring (dynamic coupling) in human-agent interaction [100]. Our goal was to study how the adaptive capabilities of an ECA, through the imitation of user's behaviors, could enhance user experience during human-agent interaction.

The setting of the experiment was an interactive installation called LoL, Laugh out Loud (ref). In this setting, a user and an ECA are listening to music inspired from the compositions by P. Shickele P.D.Q Bach. These recordings were created with the aim of making the listeners laugh. The ECA is able to tune on the fly the behavioral expressivity if its laughter according to the user's behavioral expressivity, hence creating a phenomenon of dynamic coupling between the ECA and the user. The parameters for the expressivity of ECA's laughter are the torso orientation, and the amplitude of laughter movements. For example, if the user does not laugh and

or does not move at all, ECA's laughter behavior will be inhibited. On the contrary, if the user laughs out loud and moves a lot, ECA's laughter will be amplified.

The experimental study was conducted with 32 participants. Two conditions were tested: (i) the ECA takes user's behaviors into account to modulate the expressivity of its laughter, (ii) the ECA does not take user's behaviors into account. Once the participants had listened to two short musical compositions, they had to answer a questionnaire measuring ECA's social presence. The analysis of the results revealed that when the ECA takes user's behavior into account to modulate its laughter, its social presence as perceived by the participants is greater than when it does not. The participants had the feeling that it was easier to interact with the ECA, and they had the impression they were both in the same place and that they laughed together.

In this study, the ECA behavior was generated by taking into account the user's behavior. This modulation, which takes into account human acoustic and movement features, acts upon several parameters controlling agent's animation. We chose to inhibit or to amplify the intensity of ECA's behaviors in mirroring the intensity of user's behaviors. Mirroring can be seen as a form of *alignment* between an ECA and a user. The next section presents a study exploring verbal alignment between an ECA and a user.

7.3.3 Enhancing User Engagement through Verbal Alignment by the Agent

A model of verbal alignment allowing an ECA to share appreciations with a user [30], referred as the *Appreciation Module*, was integrated in the GRETA platform as an Intention Planner. It takes as inputs ECA's preferences encoded in the *Agent Mind*. The *Appreciation Module* provides supplementary functionalities for dialogue management, in addition to the DISCO dialogue manager [116], integrated in the GRETA platform.

The development of the *Appreciation Module* is conducted in the framework of the national French project A1:1. The project's goal is to set a life-sized ECA in a museum where it plays the role of a visitor. The module, in particular, aims at enabling the ECA to *engage* museum visitors by sharing appreciations on different topics, such as an artwork, or a specific painting style. Expressing evaluation opinion or judgment is a basic activity for visitors in a museum [127], and it is important to build *rapport* and *affiliation* between two speakers, which contributes to their *engagement* [144]. Our model is twofold: it focuses on *how* an ECA can generate appreciation sentences, and *when* the ECA should effectively use them.

We modeled two types of *alignment processes* occurring during the sharing of appreciations between an user and the ECA: alignment at the lexical level through *other-repetition* (OR), and alignment at the level of *polarity* between a user's appreciation and an ECA's appreciation on the same topic. OR is the intentional repetition

Fig. 7.4 A user interacting
with Leonard during our
study conducted in the
laboratory

by the hearer of part of what the speaker has just said, in order to convey a communicative function that was not present in the first instance [6, 10, 104, 133], such as an emotional stance [131].

Our computational model enables an ECA to express emotional stances with other-repetitions [30]. This model is grounded on a previous analysis on the SEMAINE corpus [29], where we found several occurrences of ORs expressing emotional stances. Our model integrates 3 *emotional stances*: surprise, positive appreciation, negative appreciation. The selected emotional stance depends on user's appreciation and ECA's preferences, and they are expressed by the ECA in the form of a verbal *appreciation*, as defined in [82] (e.g.: "I consider it beautiful.").

An evaluation of the appreciation sentences generated by the model was conducted with the ECA Leonard designed for the A1:1 project. We simulated a small museum in our laboratory; We hung 4 pictures of existing artworks in the corridor. Each participant was asked to watch them, then to talk to Leonard (see Fig. 7.4), and ultimately fill in a questionnaire.

Thirty-four participants took part in the experiment. The results issued from subjective reports showed that the presence or absence of ORs in ECA's appreciations does not seem to have an effect on the perception of user's own engagement, or on ECA's believability perceived by the user. However, the presence of ORs in ECA's appreciations had a positive effect on participants' feeling that they shared the same appreciations as the ECA.

To improve these results, we developed an extension of the previous model dedicated at deciding *when* to trigger an exchange of appreciations between the ECA and the user [31]. This sharing of appreciations, represented as a task, is added on the fly in the dialogue plan when the user shows a low level of engagement while interacting with an ECA. For future work, we plan to conduct an evaluation of the model with different conversational strategies, such as triggering a sharing of appreciations when user engagement is high versus low.

7.3.4 Engaging Users in Multiparty Group Interaction with the Expression of Interpersonal Attitudes

Simulating group interactions and expressing social attitudes among participants can be hard to achieve. The expression of the agent's interpersonal attitude with multimodal behavior in user-agent face-to-face interactions is supported by the Greta platform [111]. However, moving to a more complex multi-party group interaction required a more powerful framework that integrated the Greta platform with the Impulsion AI Engine [101]. This latter engine combined a number of reactive social behaviors, including those reflecting Hall's personal space theory [62] and Kendon's F-formation systems [69], in a general steering framework inspired by Reynolds and colleagues [115]. The engine supports the generation of agents' reactive behaviors to make them aware of the user's presence (for example in avatar based interactions), so that users feel engaged in interaction with an agent or a group.

Impulsion's complete management of position and orientation was used in conjunction with Greta's behavior planner for facial expressions and gestures generation, in order to produce believable and dynamic group formations expressing different interpersonal attitudes both among other members of a group of agents, thus named *in-group* attitude, and towards the user, therefore named *out-group* attitude [113].

Interpersonal attitudes shape how one person relates to another person [5, p.85], in particular affiliation, according to Argyle's status and affiliation model [5, p.86], indicates the degree of liking or friendliness, ranging from unfriendly to friendly, towards another person. In the context of engagement, expressing high affiliation (i.e. friendliness) represents a valuable means for showing interest into interacting with one person.

In the context of user-agents interaction within a 3D serious game environment, the effects of both *in* and *out-group* attitude (affiliation dimension) on user's presence evaluations of a group of four agents and user's proxemics behavior in the 3D environment were studied. In two separate trials subjects had to complete the task of (1) joining a group of four agents, composed by two males and two females, and (2) reaching a point behind the group of agents with their own avatar in third person view (Fig. 7.5).

The different levels of attitude were obtained by exhibiting, for example in a friendly out-group case, smiling behavior, gazing more at the user (compared to the unfriendly case) and opening (i.e. making physical space) when the user's avatar was within the social distance of the group, according to Hall's areas [62]. The in-group attitude levels were obtained by changes of voice volume, gestures amplitude and speed, proximity among the agents, number of gaze at others, smiling behavior and turn duration.

In conclusion, results indicated that expressing interpersonal attitudes in multi-party group interaction had an impact on the evaluation of agent's presence assessed by users when those attitudes were expressed towards the user (out-group) regardless of the attitude expressed among the agents (in-group). The social presence (including engagement level) of a group of agents is dramatically reduced when an unfriendly

Fig. 7.5 The image shows a screen shot of the 3D environment as seen by the user in third person view with the avatar walking towards the group of agents in one of the conditions within the study on interpersonal attitudes in multi-party group interaction

attitude is expressed towards users. Interestingly, users in the first task (i.e. join a group) chose to get closer to those groups having unfriendly out-group attitude, possibly due to the lack of openness exhibited by the group, thus users pushed more their avatar in order to obtain a reaction. Whereas in the second task (i.e. reach destination behind the group) users walked through those having both in and out-group unfriendly attitude. This was possibly due to the bigger interpersonal space among the agents.

7.4   Conclusion and Perspectives

Considering engagement in human-agent interactions is a promising way for addressing the scientific challenges involved in generating fluent social interactions between the users and the agents. Research has unraveled many aspects concerning the detection and the generation of emotional behaviors but considering social attitudes is still an emerging topic. Existing socio-emotional interaction strategies pay more and more attention to fostering user engagement not only within a single interaction but also over several interactions. The success of socio-emotional interaction strategies can be thus evaluated by focusing on user engagement, and the present chapter provided a view of different methodologies that are used for this evaluation, from subjective tests to automatic measurements.

    The work carried out around the Greta/VIB platform takes a step in this direction by providing subjective assessments of user engagement aiming to evaluate the interaction strategies (alignment and dynamic coupling) and the expression of interpersonal attitudes in multi-party interactions. Current work on this platform concerns the integration of a system able to detect user's likes and dislikes; the development of further interaction strategies such as politeness strategies and the management of turn-taking between the user and the agent. We hope such integration will contribute to provide more fluent interactions and improve the user's engagement.

References

1. Acosta JC, Ward NG (2011) Achieving rapport with turn-by-turn, user-responsive emotional coloring. Speech Commun 53(910):1137–1148
2. Adelswärd V (1989) Laughter and dialogue: the social significance of laughter in institutional discourse. Nordic J Linguist 12(02):107–136
3. Albrecht I, Schröder M, Haber J, Seidel H (2005) Mixed feelings: expression of non-basic emotions in a muscle-based talking head. Virtual Real 8(4):201–212
4. Andre E, Rehm M, Minker W, Bühler D (2004) Endowing spoken language dialogue systems with emotional intelligence. Affective dialogue systems. Springer, Berlin, pp 178–187
5. Argyle M (1988) Bodily communication, 2nd edn. Methuen, New York
6. Bazzanella C (2011) Redundancy, repetition, and intensity in discourse. Lang Sci 33(2):243–254
7. Bee N, André E, Tober S (2009) Breaking the ice in human-agent communication: eye-gaze based initiation of contact with an embodied conversational agent. In: Proceedings of 9th international conference on Intelligent virtual agents, IVA, Amsterdam, The Netherlands, Sept 14–16, 2009, pp 229–242
8. Bee N, Pollock C, André E, Walker M (2010) Bossy or wimpy: expressing social dominance by combining gaze and linguistic behaviors. In: Proceedings of the 10th international conference on intelligent virtual agents, Springer, Berlin, Heidelberg, IVA'10, pp 265–271
9. Bergmann K, Eyssel F, Kopp S (2012) A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In: Nakano Y, Neff M, Paiva A, Walker M (eds) Intelligent virtual agents, Lecture Notes in Computer Science, vol 7502, Springer, Berlin Heidelberg, pp 126–138, doi:10.1007/978-3-642-33197-8_13
10. Bertrand R, Ferré G, Guardiola M et al (2013) French face-to-face interaction: repetition as a multimodal resource. Interaction, Coverbal Synchrony in Human-Machine, p 141
11. Beskow J (1997) Animation of talking agents. In: Benoit C, Campbell R (eds) Proceedings of the ESCA workshop on audio-visual speech processing. Rhodes, Greece, pp 149–152
12. Bickmore T, Giorgino T (2006) Health dialog systems for patients and consumers. J biomed inf 39(5):556–571
13. Bickmore T, Gruber A, Picard R (2005) Establishing the computer patient working alliance in automated health behavior change interventions. Patient Educ Couns 59(1):21–30. doi:10.1016/j.pec.2004.09.008
14. Bickmore T, Schulman D, Yin L (2010) Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 24(6):648–666
15. Bickmore T, Pfeifer L, Schulman D (2011) Relational agents improve engagement and learning in science museum visitors. Intelligent virtual agents. Springer, Berlin, pp 55–67
16. Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 12(2):293–327
17. Bohus D, Horvitz E (2014) Managing human-robot engagement with forecasts and... um... hesitations. In: Proceedings of the 16th international conference on multimodal interaction, ACM, pp 2–9
18. Boukricha H, Wachsmuth I, Hofstätter A, Grammer K (2009) Pleasure-arousal-dominance driven facial expression simulation. In: Proceedings of third international conference and

workshops on Affective computing and intelligent interaction, ACII, Amsterdam, The Netherlands, Sept 10–12 2009, pp 1–7

19. Branigan HP, Pickering MJ, Cleland AA (2000) Syntactic co-ordination in dialogue. Cognition 75(2):B13–B25

20. Brennan SE, Clark HH (1996) Conceptual pacts and lexical choice in conversation. J Exp Psychol: Learn, Mem, Cognit 22(6):1482

21. Bui TD, Heylen D, Poel M, Nijholt A (2001) Generation of facial expressions from emotion using a fuzzy rule based system. In: Stumptner M, Corbett D, Brooks M (eds) Proceedings of 14th Australian joint conference on artificial intelligence (AI 2001). Springer, Adelaide, Australia, pp 83–94

22. Burgoon J, Buller D, Hale J, de Turck M (1984) Relational messages associated with nonverbal behaviors. Hum Commun Res 10(3):351–378

23. Buschmeier H, Bergmann K, Kopp S (2009) An alignment-capable microplanner for natural language generation. In: Proceedings of the 12th European workshop on natural language generation, association for computational linguistics, pp 82–89

24. Cafaro A, Vilhjálmsson HH (2015) First impressions in human-agent virtual encounters. ACM Trans Comput-Hum Interact (TOCHI, forthcoming)

25. Cafaro A, Vilhjálmsson HH, Bickmore TW, Heylen D, Schulman D (2013) First impressions in user-agent encounters: the impact of an agent's nonverbal behavior on users' relational decisions. In: Proceedings of the 2013 international conference on autonomous agents and multi-agent systems, international foundation for autonomous agents and multiagent systems, Richland, SC, AAMAS '13, pp 1201–1202. http://dl.acm.org/citation.cfm?id=2484920.2485142

26. Cafaro A, Vilhjlmsson H, Bickmore T, Heylen D, Jhannsdttir K, Valgarsson G (2012) First impressions: users judgments of virtual agents personality and interpersonal attitude in first encounters. In: Nakano Y, Neff M, Paiva A, Walker M (eds) Intelligent virtual agents, Lecture notes in computer science, vol 7502, Springer, Berlin Heidelberg, Richland, SC, pp 67–80. doi:10.1007/978-3-642-33197-8_7

27. Callejas Z, Griol D (2011) López-Cózar R (2011) Predicting user mental states in spoken dialogue systems. EURASIP J Adv Signal Proc 1:6

28. Campano S, Sabouret N (2009) A socio-emotional model of impoliteness for non-player characters. In: 3rd international conference on affective computing and intelligent interaction and workshops. ACII 2009. IEEE, pp 1–7

29. Campano S, Durand J, Clavel C (2014) Comparative analysis of verbal alignment in human-human and human-agent interactions. In: Proceedings of the 9th international conference on language resources and evaluation (LREC'14), European language resources association (ELRA)

30. Campano S, Langlet C, Glas N, Clavel C, Pelachaud C (2015) An eca expressing appreciations. In: First international workshop on engagement in human computer interaction ENHANCE'15 held in conjuction with the 6th international conference on affective computing and intelligent interaction (ACII 2015), IEEE, sept 2015 (to appear)

31. Campano S, Clavel C, Pelachaud C (2015) I like this painting too: when an eca shares appreciations to engage users. In: Proceedings of the 14th international joint conference on Autonomous agents and multiagent systems

32. Cassell J (2000) Embodied conversational agents. MIT press, Cambridge

33. Castellano G, Mancini M, Peters C, McOwan PW (2012) Expressive copying behavior for social agents: a perceptual analysis. IEEE Trans Syst, Man Cybern, Part A: Syst Hum 42(3):776–783

34. Chartrand TL, Bargh JA (1999) The chameleon effect: the perception-behavior link and social interaction. J pers soc psychol 76(6):893

35. Choi A, Melo CD, Woo W, Gratch J (2012) Affective engagement to emotional facial expressions of embodied social agents in a decision-making game. Comput Animat Virtual Worlds 23(3–4):331–342

36. Chollet M, Ochs M, Clavel C, Pelachaud C, (2013) A multimodal corpus approach to the design of virtual recruiters. In, (2013) humaine association conference on affective computing and intelligent interaction, ACII 2013. Geneva, Switzerland, Sept 2–5:19–24
37. Clavel C, Callejas Z (2015) Sentiment analysis: from opinion mining to human-agent interaction. IEEE Trans Affect Comput 99:1. doi:10.1109/TAFFC.2015.2444846
38. Clavel C, Vasilescu I, Devillers L, Richard G, Ehrette T (2008) Fear-type emotions recognition for future audio-based surveillance systems. Speech Commun 50:487–503
39. Courgeon M, Clavel C, Martin J (2014) Modeling facial signs of appraisal during interaction: impact on users' perception and behavior. In: International conference on autonomous agents and multi-agent systems, AAMAS '14, Paris, France, May 5–9 2014, pp 765–772
40. Creswell JW (2013) Research design: Qualitative, quantitative, and mixed methods approaches (Sage publications)
41. De Carolis B, Pelachaud C, Poggi I, Steedman M (2004) Apml, a markup language for believable behavior generation. In: Prendinger H, Ishizuka M (eds) Life-like characters, cognitive technologies, Springer, Berlin Heidelberg, pp 65–85. doi:10.1007/978-3-662-08373-4_4
42. Delaherche E, Chetouani M, Mahdhaoui A, Saint-Georges C, Viaux S, Cohen D (2012) Interpersonal synchrony: a survey of evaluation methods across disciplines. IEEE Trans Affect Comput 3(3):349–365
43. Dias J, Paiva A (2005) Feeling and reasoning: a computational model for emotional agents. In: Proceedings of 12th Portuguese conference on artificial intelligence, EPIA 2005, Springer, pp 127–140
44. Ding Y, Prepin K, Huang J, Pelachaud C, Artières T (2014) Laughter animation synthesis. In: Proceedings of the 2014 international conference on autonomous agents and multi-agent systems, AAMAS '14, pp 773–780
45. D'Mello S, Graesser A (2013) AutoTutor and affective autotutor: learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans Interact Intell Syst 2(4):1–39
46. D'Mello S, Olney A, Williams C, Hays P (2012) Gaze tutor: a gaze-reactive intelligent tutoring system. Int J hum-comput stud 70(5):377–398
47. Dybala P, Ptaszynski M, Rzepka R, Araki K (2009) Activating humans with humor - a dialogue system that users want to interact with. IEICE Trans Inf Syst E92-D(12):2394–2401
48. Ekman P (2003) Emotions revealed. Times Books (US), London, New York. Weidenfeld and Nicolson (world)
49. Ekman P, Friesen W (1975) Unmasking the face: a guide to recognizing emotions from facial clues. Prentice-Hall Inc, Englewood Cliffs
50. Ekman P, Friesen W, Hager J (2002) The facial action coding system, 2nd edn. Research Nexus eBook, London, Salt Lake City, Weidenfeld and Nicolson (world)
51. Fontaine JR, Scherer KR, Roesch EB, Ellsworth P (2007) The world of emotion is not two-dimensional. Psychol Sci 13:1050–1057
52. Fukayama A, Ohno T, Mukawa N, Sawaki M, Hagita N (2002) Messages embedded in gaze of interface agents–impression management with agent's gaze. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '02, pp 41–48. doi:10.1145/503376.503385
53. Garrod S, Anderson A (1987) Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. Cognition 27(2):181–218
54. Giles H, Coupland N, Coupland J (1991) Contexts of accommodation: developments in applied sociolinguistics, Cambridge, Cambridge University Press, chap 1–Accommodation theory: communication context, and consequence
55. Gillies M, Ballin D (2003) A model of interpersonal attitude and posture generation. In: Proceedings of 4th international workshop on Intelligent agents, IVA 2003, Kloster Irsee, Germany, Sept 15–17 2003, pp 88–92
56. Gillies M, Ballin D (2004) Integrating autonomous behavior and user control for believable agents. In: 3rd International joint conference on autonomous agents and multiagent systems AAMAS, New York, NY, USA, August 19–23 2004, pp 336–343

57. Glas N, Pelachaud C (2014) Politeness versus perceived engagement: an experimental study. In: Proceedings of the 11th international workshop on natural language processing and cognitive science
58. Gockley R, Bruce A, Forlizzi J, Michalowski M, Mundell A, Rosenthal S, Sellner B, Simmons R, Snipes K, Schultz A, Wang J (2005) Designing robots for long-term social interaction. In: IEEE/RSJ International conference on intelligent robots and systems (IROS), pp 1338–1343. doi:10.1109/IROS.2005.1545303
59. Grammer K, Oberzaucher E (2006) The reconstruction of facial expressions in embodied systems. ZiF: mitteilungen, vol 2
60. Gratch J, Kang SH, Wang N (2013) Using social agents to explore theories of rapport and emotional resonance. Oxford University Press, Oxford, Social emotions in nature and artifact, p 181
61. Griol D, Molina JM, Callejas Z (2014) Modeling the user state for context-aware spoken interaction in ambient assisted living. Appl Intell 40:1–23
62. Hall ET (1966) The hidden dimension. Doubleday, Garden City
63. Hall L, Woods S, Aylett R, Newall L, Paiva A (2005) Achieving empathic engagement through affective interaction with synthetic characters. Affective computing and intelligent interaction. Springer, Heidelberg, pp 731–738
64. Hess U, Philippot P, Blairy S (1999) The social context of nonverbal behavior mimicry. cambridge University Press, Cambridge, p 213
65. Heylen D, Kopp S, Marsella SC, Pelachaud C, Vilhjálmsson HH (2008) The next step towards a function markup language. Proceedings of the 8th international conference on Intelligent Virtual Agents. Springer, Berlin, IVA, pp 270–280
66. Huber T, Ruch W (2007) Laughter as a uniform category? a historic analysis of different types of laughter. In: 10th Congress of the Swiss Society of Psychology. University of Zurich, Switzerland
67. Ivaldi S, Anzalone SM, Rousseau W, Sigaud O, Chetouani M (2014) Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. Front Neurorobot 8:5
68. Izard C (1994) Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychol Bull 115:288–299
69. Kendon A (1990) Conducting interaction: patterns of behavior in focused encounters (Studies in interactional sociolinguistics). Cambridge University Press, New York
70. Kopp S, Krenn B, Marsella S, Marshall AN, Pelachaud C, Pirker H, Thórisson KR, Vilhjálmsson HH (2006) Towards a common framework for multimodal generation: the behavior markup language. Proceedings of the 6th international conference on Intelligent virtual agents. Springer, Heidelberg, IVA, pp 205–217
71. Kopp S (2010) Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. Speech Commun 52(6):587–597
72. Kopp S, Gesellensetter L, Krämer NC, Wachsmuth I (2005) A conversational agent as museum guide design and evaluation of a real-world application. In: Panayiotopoulos T, Gratch J, Aylett R, Ballin D, Olivier P, Rist T (eds) Intelligent virtual agents, Lecture notes in computer science, vol 3661, Springer, Berlin, pp 329–343. doi:10.1007/11550617_28
73. Lambertz K (2011) Back-channelling: The use of yeah and mm to portray engaged listenership. Griffith working papers in pragmatics and intercultural communication. vol 4, pp 11–18
74. Langlet C, Clavel C (2014) Modelling user's attitudinal reactions to the agent utterances: focus on the verbal content. In: Proceedings of 5th international workshop on corpora for research on emotion, sentiment & social signals (ES3), Reykjavik, Iceland
75. Langlet C, Clavel C (2015) Improving social relationships in face-to-face human-agent interactions: when the agent wants to know users likes and dislikes. In: Proceedings of annual meeting of the association for computational linguistic
76. Lee J, Marsella S (2006) Nonverbal behavior generator for embodied conversational agents. In: Gratch J, Young M, Aylett R, Ballin D, Olivier P (eds), Proceedings 6th international conference intelligent virtual agents (IVA) Marina Del Rey, CA Springer, LNCS, vol 4133, Aug 21–23 2006, pp 243–255

77. Litman DJ, Forbes-Riley K (2006) Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. Speech Commun 48(5):559–590

78. Lombard M, Weinstein L, Ditton T (2011) Measuring telepresence: the validity of the temple presence inventory (tpi) in a gaming context. In: Proceedings of 2011 Annual conference of the International Society for Presence Research (ISPR), ISPR 2011

79. Maat M, Heylen D (2009) Turn management or impression management? In: Proceedings of the 9th international conference on intelligent virtual agents, Springer, Heidelberg, IVA, pp 467–473. doi:10.1007/978-3-642-04380-2_51

80. Maat MT, Truong KP, Heylen D (2010) How turn-taking strategies influence users' impressions of an agent. In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A (eds) Intelligent virtual agents, Springer, Heidelberg, Lecture notes in computer science, vol 6356, pp 441–453. doi:10.1007/978-3-642-15892-6_48

81. Magnusson M (2000) Discovering hidden time patterns in behavior: T-patterns and their detection. Behav Res Meth, Instrum Comput 32(1):93–110

82. Martin JR, White PR (2005) The language of evaluation. Palgrave Macmillan Basingstoke, New York

83. Martin JR, White PR (2005) The Language of Evaluation. Appraisal in English, Palgrave Macmillan Basingstoke, New York

84. Mehrabian A (1980) Basic dimensions for a general psychological theory: implications for personality, social, environmental, and developmental studies. Oelgeschlager, Gunn & Hain, Cambridge

85. Mol L, Krahmer E, Maes A, Swerts M (2012) Adaptation in gesture: converging hands or converging minds? J Mem Lang 66(1):249–264

86. Morency LP, de Kok I, Gratch J (2010) A probabilistic multimodal approach for predicting listener backchannels. Auto Agents Multi-Agent Syst 20(1):70–84

87. Munezero MD, Suero Montero C, Sutinen E, Pajunen J (2014) Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. IEEE Trans Affect Comput 5:101–111

88. Nakano YI, 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, New York, IUI, pp 139–148. doi:10.1145/1719970.1719990

89. Nicolle J, Rapp V, Bailly K, Prevost L, Chetouani M (2012) Robust continuous prediction of human emotions using multiscale dynamic cues. In: Proceedings of the 14th ACM international conference on multimodal interaction, pp 501–508

90. Niederhoffer KG, Pennebaker JW (2002) Linguistic style matching in social interaction. J Lang Soc Psychol 21(4):337–360

91. Niewiadomski R, Pelachaud C (2007) Model of facial expressions management for an embodied conversational agent. In: Proceedings of 2nd international conference on affective computing and intelligent interaction ACII, Lisbon

92. Niewiadomski R, Hyniewska SJ, Pelachaud C (2011) Constraint-based model for synthesis of multimodal sequential expressions of emotions. IEEE Trans Affect Comput 2(3):134–146

93. Niewiadomski R, Demeure V, Pelachaud C (2010) Warmth, competence, believability and virtual agents. In: Allbeck J, Badler N, Bickmore T, Pelachaud C, Safonova A (eds) Intelligent virtual agents, Lecture notes in computer science, vol 6356, Springer, Heidelberg, pp 272–285. doi:10.1007/978-3-642-15892-6_29

94. Niewiadomski R, Obaid M, Bevacqua E, Looser J, Anh LQ, Pelachaud C (2011b) Cross-media agent platform. In: Proceedings of the 16th international conference on 3D web technology, ACM, New York, Web 3D, pp 11–19. doi:10.1145/2010425.2010428

95. Ochs M, Prepin K, Pelachaud C (2013) From emotions to interpersonal stances: Multi-level analysis of smiling virtual characters. In: Proceedings of 2013 humaine association conference on affective computing and intelligent interaction (ACII), IEEE, pp 258–263

96. Osherenko A, Andre E, Vogt T (2009) Affect sensing in speech: Studying fusion of linguistic and acoustic features. In: Proceedings of affective computing and intelligent interaction (ACII)

97. Paleari M, Lisetti C (2006) Psychologically grounded avatars expressions. In: Proceedings of first workshop on emotion and computing at KI, 29th annual conference on artificial intelligence, Bremen, Germany
98. Pardo JS (2006) On phonetic convergence during conversational interaction. J Acoust Soc Am 119(4):2382–2393
99. Pecune F, Cafaro A, Chollet M, Philippe P, Pelachaud C (2014) Suggestions for extending saiba with the vib platform. In: Workshop on architectures and standards for IVAs, held at the '14th international conference on intelligent virtual agents (IVA)', Bielefeld eCollections, pp 16–20
100. Pecune F, Mancini M, Biancardi B, Varni G, Volpe G, Ding Y, Pelachaud C, Camurri A (2015) Laughing with a virtual agent. In: Proceedings of the 14th international joint conference on Autonomous agents and multiagent systems
101. Pedica C, Vilhjálmsson HH (2010) Spontaneous avatar behavior for human territoriality. Appl Artif Intell 24(6):575–593. doi:10.1080/08839514.2010.492165
102. Pekrun R, Cusack A, Murayama K, Elliot AJ, Thomas K (2014) The power of anticipated feedback: effects on students' achievement goals and achievement emotions. Learn Instruct 29:115–124
103. Pelachaud C, Badler N, Steedman M (1996) Generating facial expressions for speech. Cognit Sci 20(1):1–46
104. Perrin L, Deshaies D, Paradis C (2003) Pragmatic functions of local diaphonic repetitions in conversation. J Pragmat 35(12):1843–1860
105. Peters C, Castellano G, de Freitas S (2009) An exploration of user engagement in hci. In: Proceedings of the international workshop on affective-aware virtual agents and social robots, ACM, p 9
106. Pickering MJ, Garrod S (2004) Toward a mechanistic psychology of dialogue. Behav brain sci 27(02):169–190
107. Plutchnik R (1980) Emotion: a psychoevolutionary synthesis. Harper and Row, NY
108. Poggi I (2007) Mind, hands, face and body: a goal and belief view of multimodal communication. Weidler, Berlin
109. Prepin K, Ochs M, Pelachaud C (2013) Beyond backchannels: co-construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents. In: International conference CogSci (Annual conference of the cognitive science society)
110. Provine RR (2001) Laughter: a scientific investigation. Penguin, New York
111. Ravenet B, Ochs M, Pelachaud C (2013) From a user-created corpus of virtual agents non-verbal behavior to a computational model of interpersonal attitudes. In: Aylett R, Krenn B, Pelachaud C, Shimodaira H (eds) Intelligent virtual agents, Lecture notes in computer science. Springer, Heidelberg, pp 263–274. doi:10.1007/978-3-642-40415-3_23
112. Ravenet B, Ochs M, Pelachaud C (2013) From a user-created corpus of virtual agent's non-verbal behavior to a computational model of interpersonal attitudes. In: Proceedings 13th international conference intelligent virtual agents, IVA 2013, Edinburgh, UK, August 29–31 2013, pp 263–274
113. Ravenet B, Cafaro A, Ochs M, Pelachaud C (2014) Interpersonal attitude of a speaking agent in simulated group conversations. In: Bickmore T, Marsella S, Sidner C (eds) Intelligent virtual agents, Lecture notes in computer science, vol 8637, Springer International Publishing, pp 345–349. doi:10.1007/978-3-319-09767-1_45
114. Ravenet B, Cafaro A, Ochs M, Pelachaud C (2014) Interpersonal attitude of a speaking agent in simulated group conversations. In: Proceedings of 14th international conference on intelligent virtual agents, IVA 2014, Boston, August 27–29, 2014, pp 345–349
115. Reynolds C (1999) Steering behaviors for autonomous characters. Miller Freeman Game Groups, San Francisco, Proceedings of the game developers conference, p 763
116. Rich C, Sidner CL (2012) Using collaborative discourse theory to partially automate dialogue tree authoring. In: Proceedings of the 12th international conference on intelligent virtual agents, Springer, Heidelberg, IVA'12, pp 327–340. doi:10.1007/978-3-642-33197-8_34
117. Russell J (1980) A circumplex model of affect. J Person Soc Psychol 39:1161–1178

118. Ruttkay Z, Noot H, ten Hagen P (2003) Emotion disc and emotion squares: tools to explore the facial expression face. Comput Graph Forum 22(1):49–53

119. Scherer K (2000) Emotion. In: Stroebe W, Hewstone M (eds) Introduction to social psychology, a European perspective. Oxford Blackwell Publishers, Oxford, pp 151–191

120. Scherer K (2005) What are emotions? And how can they be measured? Soc sci inf 44(4):695–729

121. Schmitt A, Polzehl T, Minker W (2010) Facing reality: simulating deployment of anger recognition in IVR systems. In: Lee GG, Mariani J, Minker W, Nakamura S (eds) Spoken dialogue systems for ambient environments. Lecture notes in computer scienceSpringer, Heidelberg, pp 122–131

122. Schroder M, Bevacqua E, Cowie R, Eyben F, Gunes H, Heylen D, Ter Maat M, McKeown G, Pammi S, Pantic M et al (2012) Building autonomous sensitive artificial listeners. IEEE Trans Affect Comput 3(2):165–183

123. Schuller B, Rigoll G, Lang M (2004) Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. Proceedings of ICASSP, Montreal 1:577–580

124. Schuller B, Batliner A, Steidl S, Seppi D (2011) Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. Speech Commun 53(9):1062–1087

125. Sidner CL, Dzikovska M (2002) Human-robot interaction: engagement between humans and robots for hosting activities. In: Proceedings of the 4th IEEE international conference on multimodal interfaces, IEEE Computer Society, p 123

126. Sidner CL, Lee C, Kidd CD, Lesh N, Rich C (2005) Explorations in engagement for humans and robots. Artif Intell 166(1):140–164

127. Silverman LH (1999) Meaning making matters: communication, consequences, and exhibit design. Exhibitionist

128. Smith C, Crook N, Dobnik S, Charlton D (2011) Interaction strategies for an affective conversational agent. Presence: teleoperators and virtual environments. MIT Press, Cambridge, pp 395–411

129. Srikant R, Agrawal R (1996) Mining sequential patterns: generalizations and performance improvements. Adv Database Technol 1057:1–17

130. Stivers T (2008) Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. Res Lang soc interact 41(1):31–57

131. Svennevig J (2004) Other-repetition as display of hearing, understanding and emotional stance. Discourse studies 6(4):489–516

132. Takashima K, Omori Y, Yoshimoto Y, Itoh Y, Kitamura Y, Kishino F (2008) Effects of avatar's blinking animation on person impressions. In: Proceedings of graphics interface, Toronto, Ontario, Canada, GI '08, Canadian Information Processing Society, pp 169–176. doi:10.1145/1375714.1375744

133. Tannen D (1992) Talking voices: repetition, dialogue, and imagery in conversational discourse. cambridge University Press, Cambridge

134. Traum D, Aggarwal P, Artstein R, Foutz S, Gerten J, Katsamanis A, Leuski A, Noren D, Swartout W (2012) Ada and grace: Direct interaction with museum visitors. In: Nakano Y, Neff M, Paiva A, Walker M (eds) Intelligent virtual agents, Lecture notes in computer science, vol 7502, Berlin Heidelberg, Springer, pp 245–251. doi:10.1007/978-3-642-33197-8_25

135. Truong KP, Poppe R, Heylen D (2010) A rule-based backchannel prediction model using pitch and pause information, pp 3058–3061

136. Tsapatsoulis N, Raouzaiou A, Kollias S, Cowie R, Douglas-Cowie E (2002) Emotion recognition and synthesis based on MPEG-4 FAPs in MPEG-4 facial animation. In: Pandzic IS, Forcheimer R (eds) MPEG4 facial animation–the standard, implementations and applications. Wiley, New York

137. Vilhjálmsson HH, Cantelmo N, Cassell J, E Chafai N, Kipp M, Kopp S, Mancini M, Marsella S, Marshall AN, Pelachaud C, Ruttkay Z, Thórisson KR, Welbergen H, Werf RJ (2007) The behavior markup language: recent developments and challenges. In: Proceedings of the

7th international conference on intelligent virtual agents, Heidelberg, IVA '07, Springer, pp 99–111

138. Vinciarelli A, Pantic M, Bourlard H (2009) Social signal processing: survey of an emerging domain. Image Vis Comput 27(12):1743–1759
139. Wang N, Johnson WL, Mayer RE, Rizzo P, Shaw E, Collins H (2008) The politeness effect: pedagogical agents and learning outcomes. Int J Hum-Comput Studies 66(2):98–112
140. Ward A, Litman D (2007) Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In: Proceedings of the SLaTE Workshop on speech and language technology in education
141. With S, Kaiser S (2011) Sequential patterning of facial actions in the production and perception of emotional expressions. Swiss Journal of Psychology / Schweizerische Zeitschrift fr Psychologie / Revue Suisse de Psychologie 70(4):241–252
142. Yildirim S, Narayanan S, Potamianos A (2011) Detecting emotional state of a child in a conversational computer game. Comput Speech Lang 25(1):29–44
143. Zhang L (2009) Exploration of affect sensing from speech and metaphorical text. In: learning by playing. Game-based education system design and development, springer, Berlin, pp 251–262
144. Zhao R, Papangelis A, Cassell J (2014) Towards a dyadic computational model of rapport management for human-virtual agent interaction. Intelligent virtual agents. Springer, Berlin, pp 514–527