A user-perception based approach to create smiling embodied conversational agents

Ochs, M., Pelachaud, C., & McKeown, G. (2017). A user-perception based approach to create smiling embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems (TiiS), 7*(1), [4]. https://doi.org/10.1145/2925993

Document Version:
Peer reviewed version

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

A user-perception based approach to create smiling embodied conversational agents

Magalie Ochs, Aix Marseille Université, CNRS, ENSAM, Université de Toulon, LSIS, Marseille, France
Catherine Pelachaud, CNRS-LTCI Télécom ParisTech, Multimedia group, Paris, France
Gary McKeown, Queen's University of Belfast, School of Psychology, Northern Ireland, UK

In order to improve the social capabilities of embodied conversational agents, we propose a computational model to enable agents to automatically select and display appropriate smiling behavior during human-machine interaction. A smile may convey different communicative intentions depending on subtle characteristics of the facial expression and contextual cues. So, to construct such a model, as a first step, we explore the morphological and dynamic characteristics of different types of smile (polite, amused and embarrassed smiles) that an embodied conversational agent may display. The resulting lexicon of smiles is based on a corpus of virtual agent's smiles directly created by users and analyzed through a machine learning technique. Moreover, during an interaction, the expression of smile impacts on the observer's perception of the interpersonal stance of the speaker. As a second step, we propose a probabilistic model to automatically compute the user's potential perception of the embodied conversational agent's social stance depending on its smiling behavior and on its physical appearance. This model, based on a corpus of users' perception of smiling and non-smiling virtual agents, enables a virtual agent to determine the appropriate smiling behavior to adopt given the interpersonal stance it wants to express. An experiment using real human-virtual agent interaction provided some validation of the proposed model.

Additional Key Words and Phrases: H.5.2 User Interfaces, Embodied Conversational Agent; Smiles; Human-Machine Interaction

1. INTRODUCTION

Computers are increasingly used in roles that are typically fulfilled by humans, such as virtual tutors in a learning class or virtual assistants for task realization. When computers are used in these roles they are often embodied by animated cartoon or human like virtual characters, called *Embodied Conversational Agents* (ECA) [Cassell 2000]. This enables a more natural style of communication for the human and allows the computer to avail of both verbal and non-verbal behavior channels of communication. Several studies have demonstrated the acceptance and the efficiency of such agents [Burgoon et al. 2016; Dehn and van Mulken 2000; Krämer 2008]; indeed, the *persona effect* reveals that the presence of an ECA improves the experience of an interaction for the user (for instance [Mayer and DaPra 2012; Pardo et al. 2009]). Moreover, when people interact with such virtual agents, they tend to react naturally and socially as they would do with another person [Wang et al. 2005; Nowak and Biocca 2003; Krämer 2008].

These kinds of human-machine interaction are facilitated by endowing ECAs with social capabilities. A particularly important social cue during social interactions is the *smile* [Knapp et al. 2013]. Smiling is an important social signal in negotiating interpersonal relationships. Smiling individuals are seen as more relaxed, kind, warm, attractive, successful, sociable, polite, happy, honest, having more of a sense of humor, and being less dominant [Bernstein et al. 2010; Ketelaar et al. 2012; Otta et al. 1996; ODoherty et al. 2003]. Smiling—often in combination with laughter—plays an important role in establishing affiliation between conversational partners and establishing social bonds that engage a polite interpersonal environment; this is likely to make an interlocutor more tolerant, more likely to pursue a smooth non-aggressive interaction, and more prepared to repair a conversation when it breaks down [Glenn 2003]. For these reasons smiling is a worthwhile addition to the repertoire of social signals available to an ECA. However there are risks too, smiling or laughing at the wrong time or

in an inappropriate social context could violate social norms and have a negative effect on an interaction [McKeown et al. 2015].

Subtle characteristics of a smile facial expression can influence the way in which the smile's meaning is interpreted. For instance, a smile may communicate amusement, politeness, or embarrassment depending on which facial muscles are activated. People consciously or unconsciously display these different smiles during an interaction and are able to distinguish them when they are expressed by their interlocutor [Frank et al. 1993]. The smile social signal is "profoundly influential" [Knapp et al. 2013]; a smile expression can have a strong impact on how an individual is perceived by another person. In particular, *smiling behavior*, that is, when a smile occurs and which type of smile is expressed during an interaction, may determine the perceived *interpersonal stances* [Deutsch et al. 1987; Edinger and Patterson 1983; Lau 1982; Moore 1985; Reis et al. 1990; Bernstein et al. 2010; Ketelaar et al. 2012; Otta et al. 1996; ODoherty et al. 2003]. An *interpersonal stance* corresponds to an attitude, spontaneously or strategically expressed, that conveys the relationship of a person to the interlocutor (for example "warm" or "polite") [Kielsing 2009; Scherer 2005]. The effect of any given smile behavior might vary from positive to negative depending on the facial expression characteristics of smile, when it is expressed, and in response to what—the social context. The social context corresponds to the environment where the agent is with different factors influencing its behavior : the situational context (e.g. social occasion), the social role, the cultural conventions, and the social norms. [Riek and Robinson 2011].

In order to improve the social capabilities of ECAs, this paper proposes a computational model designed to enable embodied agents to select and display appropriate *smiling behavior* during a human-machine interaction. To create such smiling agents, two major issues must be addressed: agents need to be able to produce appropriate smiles, and they must know when in a conversation it is appropriate to produce them. To address the first issue—endowing an agent with the capability to express different types of smile—the key morphological and dynamic characteristics relevant to smiling need to be identified in order to adequately animate the agent's face (Section 2.1). For this purpose, we have constructed a lexicon of smiling facial expressions containing different types of smile that convey different meanings, such as amusement or politeness (Section 4). The second issue—enabling an ECA to select its smiles depending on the potential effect of its expressions on the way users perceive its interpersonal stances—is more complex. Such an agent should be able to decide when and which smiles to express given a stance that it wants to convey to a user (Section 2.2). To address this issue, we have constructed a model to automatically compute, in real-time during an interaction, the effects of different smiling behaviors on how the user will perceive the ECAs' interpersonal stances (Section 5).

A *user-perception* methodology—combined with machine learning techniques—was used to create the ECA's computational model of smiling behaviors. The lexicon of smiles and the model that infers the effects of the ECA's smiles on the perception of stances, are derived from user perceptions. More precisely, potential users actively participated in the creation of the computational models of smiling behaviors, both to define the characteristics of different smile types and to estimate the effects of smiling behavior on the perceived interpersonal stances. Machine learning techniques were used to extract the information from the data collected from users.

The paper is organized in 4 sections. The first section (Section 2) reviews the literature on smiles with particular respect to the morphological and dynamic characteris-

tics of human and virtual agent smiles (Section 2.1) and the effects of smiling people or virtual agents on other people's perceptions (Section 2.2). Section 3 addresses the issues concerning the integration of a computational model of smiles into an ECA. In Section 4 we detail the creation of the lexicon of facial expressions conveying different types of smile. Section 5 is dedicated to the model of the effects of the smiling behavior of ECAs on the user perceptions of interpersonal stances. In concluding we discuss the limits of the presented work and the next steps.

2. SMILES IN HUMAN-HUMAN AND HUMAN-MACHINE INTERACTION: EMPIRICAL AND THEORETICAL BACKGROUND

2.1. Morphological and dynamic characteristics of smiles

2.1.1. Human's smiles. A smile is one of the simplest and most easily recognized facial expressions [Ekman and Friesen 1982]. To create a smile, the two zygomatic major muscles, on either side of the face, have to be activated. However, other muscles may be involved in a smile expression. Depending on which muscles are activated and how they are activated, smiles with different meanings can be distinguished. Ekman [Ekman 2009] identified 18 smile types and proposed that there might be as many as 50 in all. The most common one is the *amused smile*, also called a felt smile, Duchenne smile, enjoyment smile, or genuine smile. The amused smile is often opposed to the *polite smile*, also known as a non-Duchenne smile, false smile, social smile, masking smile, or controlled smile [Frank et al. 1993]. Perceptual studies [Frank et al. 1993] have shown that people consciously and unconsciously distinguish between an amusement smile and a polite one. Smiles are typically associated with positive aspects of interaction, but they may also occur in negative situations. For instance, a specific smile appears in the facial expression of embarrassment [Keltner 1995], anxiety [Harrigan and O'Connell 1996], or frustration [Hoque et al. 2012].

The issue of the existence of "types" of emotions and corresponding facial expressions remains, of course, a contentious issue within the psychological literature. Facial expressions are continuous in nature and not marked by abrupt step changes from one state to another, but gradual, although often rapid, movements between states–this becomes more apparent when dealing with dynamic and spontaneous facial expression stimuli rather than static photo-based expression stimuli. There are also dangers in starting with the linguistic descriptions and moving towards perception-based imagery; the choice of words inevitably has an effect on the perception of the expression. Word choice can reify differences in perceptions which can have an impact on the materials in a study. This argument is perhaps easiest to appreciate in the argument over the use of acted and spontaneous expressions of emotion. In an acted situation an actor is asked to express several emotions defined by words that are emotional labels (e.g. happy, sad, angry), the actor then does his or her best to create three emotional expressions that maximally discriminate between these emotions–if they did not create obviously discriminable expressions they would not be doing their job correctly. However, this sets up a confound in an experimental process when people or machines are later asked to categorize such expressions as they have been devised, by the actor, to be maximally easy to discriminate. One way to get around this situation is to use a spontaneous emotion induction scenario is to devise a task that may produce the desired emotion and associated expressions, but give no explicit instruction to the participant to do so–this produces more ecologically valid expressions [Sneddon et al. 2012]. In a more subtle version of this argument asking people to categorize stimuli on the basis of words sets up preconceptions and expectations of the perceptions they are looking for, this carries the risk that stereotypical perceptions are induced rather than

a possible range of continuous perceptions associated with the phenomena under investigation. In this paper we have opted for a user perception methodology to actively involve the participants in the creation of the models of smiling behavior. However, as we cannot make agents spontaneously generate smiles in response to a task we anchor this methodology using three "types" of smile–this runs the risk of inducing more stereotypical perceptions and consequently the creation of more stereotypical smile stimuli.

In this paper, the three smile types we focus on are: *amused*, *polite*, and *embarrassed*. These smiles have been explored from both the *encoder* point of view (*i.e.* the person who smiles) [Ekman and Friesen 1982; Krumhuber and Manstead 2009] and from the *decoder* point of view (*i.e.* the person who perceives the smile) [Ambadar et al. 2009; Krumhuber and Manstead 2009].

Polite, amused, and embarrassed smiles can be distinguished through their differing morphological and dynamic characteristics. *Morphological characteristics* are structural changes in the face such as the mouth opening or cheek raising, whereas *dynamic characteristics* correspond to the temporal unfolding of the smile. The dynamics of facial expressions are commonly defined by three time intervals. The *onset* corresponds to the interval of time in which the expression reaches its maximal intensity starting from the neutral face. The *apex* is the time during which the expression maintains its maximal intensity. Finally, the *offset* is the interval of time in which the expression returns from maximal intensity to the neutral expression [Ekman and Friesen 1982]. As proposed in [Hoque et al. 2011], we use the terms *rise*, *sustain* and *decay* to refer to the onset, apex and offset, since the spontaneous smile often displays a "sustained region with multiple peaks" [Hoque et al. 2011].

In the literature on smiles [Ambadar et al. 2009; Krumhuber and Manstead 2009; Ekman and Friesen 1982], the following characteristics are generally considered to distinguish the amused, polite, and embarrassed smiles[1]:

— *morphological characteristics*[2]: cheek raising (Action Unit 6), lip press (Action Unit 24), zygomatic major (Action Unit 12), symmetry of the lip corners, mouth opening (Action Unit 25), and amplitude of the smile. The cited Action Units are illustrated Figure 1;
— *dynamic characteristics*: duration of the smile and velocity of the rise and decay of the smile.

Concerning the *cheek raising*, Ekman [Ekman 2003] claims the orbicularis oculi (which refers to the Action Unit (AU) 6 in the Facial Action Coding System [Ekman et al. 2002]) is activated in an amused smile. Without it, the expression of happiness seems to be insincere [Duchenne 1990]. Recently, the role of AU6 in the smile of amusement was challenged by Krumhuber and Manstead [Krumhuber and Manstead 2009]. Their study showed that the orbicularis oculi (AU6) can occur in both spontaneous and deliberate smiles. However, the perception of amused and polite smiles depends on the presence of AU6. According to Ekman [Ekman 2003], *asymmetry* is an indicator of voluntary and non-spontaneous expression, such as the polite smile. While a lip press (AU24) is often related to the smile of embarrassment [Keltner 1995]. Krumhuber and Manstead also emphasized that dynamic smile characteristics were important for the perceiver to distinguish different smiles; dynamic cues may be even more important than the AU6 [Krumhuber and Manstead 2009]. The different

---

[1]Note that other elements of the face, such as gaze and eyebrows, influence how a smile is perceived. However, in the presented work, we focus on the influence of the smile and we do not consider the other elements of the face.

[2]the Action Units (AUs) refer to the Facial Action Coding System (FACS) proposed in [Ekman et al. 2002])
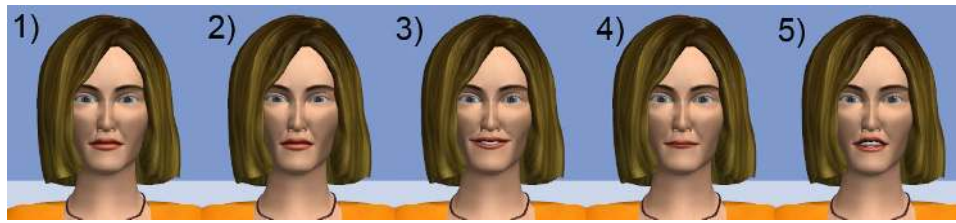
Fig. 1.   Images of an ECA displaying the different Action Units (AU) at their apex: 1) no activated AU, 2) AU6 - FACS name: Cheek Raiser - Orbicularis oculi muscle, 3) AU12 - FACS name: Lip Corner Puller, zygomaticus major muscle - 4) AU24 : FACS name: Lip Pressor - orbicularis oris muscle, 5) AU25 - FACS name: Lips Part, depressor labii inferioris, or relaxation of mentalis or orbicularis oris muscle [Ekman et al. 2002]

smile types may also have different *durations*. Felt expressions, such as the amused smile, can last from half a second to four seconds, even if the corresponding emotional state is longer [Ekman 2003; Hess and Kleck 1990]. The duration of a polite or embarrassed smile is typically shorter than 0.5 second or longer than 4 seconds [Ekman and Friesen 1982; Ekman 2003; Hess and Kleck 1990]. However, a recent study on spontaneous expression of amused and polite smiles contradicted these values showing that an amused smile may be longer than 4 seconds and that a polite smile may have a duration in the interval [2.5;4.9][3] [Hoque et al. 2011]. Not only the overall duration, but also the course of the expression is supposed to differ depending on smile type. In deliberate expressions, the rise is often abrupt or excessively short [Hoque et al. 2011; Ekman and Friesen 1982], the sustain is held too long, and the decay can be either more irregular or abrupt and short [Ekman and Friesen 1982]. However, these findings were not confirmed in [Hoque et al. 2011] who showed that amused smiles have a longer sustain period than polite smiles.

No strong consensus exists concerning the morphological and dynamic characteristics of amused, polite and embarrassed smiles. Indeed, there is a disagreement on the features of smiles, in particular concerning the cheek raising (AU6) and the dynamics of the polite and amused smiles. We then suppose that within each smile type, there are multiple smile expressions (with differences in morphological and dynamic characteristics) which express the same meaning. In the study presented in this article, we explore the different smiles that may convey an amused, polite, or embarrassed meaning. We start from the smile characteristics highlighted in the human and social sciences studies presented above to establish the relevant variables that should be used to define ECA smile expressions. In Section 4.1, we present a tool that enables users to create virtual agent's smiles by manipulating these variables.

2.1.2. Smiling virtual agents. There are several existing ECAs that *smile* during an interaction. Most of them use smiles to express a positive emotion or a positive mood [Poggi and Pelachaud 2000]. Some ECAs smile to express a specific communicative intention. For instance, in [Cassell et al. 2001], smiles are used for greetings. In [Krämer et al. 2013; Bevacqua et al. 2010a], smiles are displayed as backchannels to express understanding and liking during periods in which the user speaks. In [Theonas et al. 2008], the expression of smiles are used to create a global friendly atmosphere. In these examples and generally within the domain of virtual agents, expressed smiles typically exhibit *amused smile* characteristics as described in the previous section (Section 2.1.1).

---

[3]Considering non-shared polite smiles.

Only a few researchers have considered using different smiles in virtual agents to increase the repertoire of available facial expressions. For instance, in [Tanguy 2006], two smile types, *amused* and *polite*, are displayed by an ECA. The *amused smile* reflects an emotional state of happiness whereas the *polite smile*—called a "fake smile" in [Tanguy 2006]—is supposed to mask sadness with a smile. The characteristics of the smiles are based on the theoretical descriptions of smiles proposed in [Ekman 1992]. The amused smile is represented by lip corners raised, lower eyelids raised (Action Unit 7), and an open mouth. The polite smile is represented by an asymmetric raising of the lip corners and an expression of sadness in the upper part of the face. In [Rehm and André 2005], virtual agents mask the negative felt emotions of disgust, anger, fear, or sadness with a smile. Two types of facial expression were created according to Ekman's description [Ekman and Friesen 1975]. The first expression corresponds to the felt emotion of happiness (the *amused smile*). The second expression corresponds to the negative felt emotions (e.g. disgust) masked by unfelt happiness. The expression of unfelt happiness lacks the AU6 activity and is asymmetric; this would correspond to a *polite smile* in the terminology of this paper. Niewiadomski and Pelachaud [Niewiadomski and Pelachaud 2007] proposed an algorithm to generate complex facial expressions, such as masked or fake expressions. In particular, it is possible to generate different expressions of joy: a felt and a fake one. The *felt* expression of joy uses the reliable features (AU6) and includes an *amused smile* while the second one is asymmetric and would correspond to a *polite smile*. In a nutshell, based on the description of smiles in the literature, several agents may display a polite or an amused smile. The main limit of the existing works is that each smile type is expressed with a unique facial expression.

The descriptions of human facial expressions identified in psychology are commonly used to create repertoires of virtual agent facial expressions. As highlighted in [Grammer and Oberzaucher 2006], most studies on emotional facial expressions attribute an emotion label to a facial expression. This approach supposes that a one to one correspondence exists between an emotion and a facial expression. This one to one mapping has been questioned since each emotion may be represented by different facial expressions. Other methods have been proposed that may be more suitable to capturing the one to many correspondence between emotions and facial expressions. For instance, [Snodgrass 1992; Grammer and Oberzaucher 2006] have suggested analyzing the relationship between emotional dimensions (such as pleasure and arousal) and muscle activation (i.e. action unit). In [Grammer and Oberzaucher 2006], 2904 different randomly expressive faces have been generated and rated by more than 400 participants. A classification method was then used to categorize which muscular activition happens to convey emotions. The main problem with this method is the number of required participants for a repetitive and time-consuming task of rating each facial expression. In this article, to overcome the requirements of having participants going through repetitive tasks, we propose an alternative methodology to identify the one to many correspondences between smile types and ECA's facial expressions by asking users to directly create the facial expressions of smile types. As far as we know, no research has explored the development of virtual agent facial expressions that have been created and defined directly by users.

The originality of the work presented in this article is the *user-perception approach* used to create an ECA's smile expressions. This method avoids the traditional approach of creating a repertoire of facial expressions by asking users to label predefined facial expressions. Instead, users are placed at the heart of the facial expression creation process. There are additional advantages to this approach as it provides data that en-

able a deeper investigation of the relationship between the different facial expressions and the corresponding smile types. We present the method in details in Section 4.1.

2.2. The impact of smiles on perception

2.2.1. The impact of a smiling human on the perception of inter-personal relationships. Several studies have shown that individuals who smile are perceived more positively than non-smiling persons. Smiling people are viewed as more relaxed, kind, warm, attractive, successful, sociable, polite, happy, honest, having more of a sense of humor, and less dominant [Deutsch et al. 1987; Edinger and Patterson 1983; Lau 1982; Moore 1985; Reis et al. 1990].

In Western society, gender has a significant effect on smiles. Women smile more than men and are also expected to do so [Deutsch et al. 1987; LaFrance and Hecht 1995]. For instance, [Deutsch et al. 1987] revealed that the absence of smile can be detrimental to a woman's image in comparison to a man, whereas there is no significant difference in image perception between smiling men and women. Non-smiling women are perceived less happy and relaxed than non-smiling men. People expect that women smile more than men, and consequently, a deviation from that expected behavior negatively influences the perception of non-smiling women. No distinction between polite and amused smiles is considered in the study. Moreover, as shown in [Hess et al. 2000], the expectation that women smile more often can lead to a woman's smile being perceived as less informative in comparison to a man's smile, while men's amused smiles are perceived as more intense than those of women.

The type of displayed smile can affect an observer's perception of the smiler. People showing an amused smile are perceived as more expressive, natural, outgoing, sociable, relaxed, likeable, and pleasant than when they show polite smiles [Frank et al. 1993; LaFrance and Hecht 1995]. Amused smiling faces are also perceived as being more sociable and generous than polite smiling faces [Mehu et al. 2007].

2.2.2. The impact of smiling virtual agents on user's perception. In the domain of ECAs, several studies have explored the effects of the expression of smiles on the user's perception.

For instance, in [Krumhuber et al. 2007], virtual faces displaying an amused smile were rated as more attractive, more trustworthy, and less dominant than those showing a polite smile. In [Rehm and André 2005], a virtual agent expressing an amused smile was perceived as more reliable, trustworthy, convincing, credible, and more certain about what it said compared to the agent expressing a negative emotion masked by a smile (corresponding to a polite smile). The conditions in which smiles are expressed may also have an impact. For instance, [Theonas et al. 2008] showed that virtual agent smiles, expressed in an appropriate situation, enable the creation of a sense of comfort and warmth, and a global friendly and lively atmosphere. Smiles may also lead to smiles. For instance, [Krämer et al. 2013] showed that an ECA's smiles elicit users' smiles. Finally, [Krumhuber et al. 2007] noticed an appearance effect: smiles shown by virtual agents with female appearance were judged as less authentic than those displayed by agents with male appearance, regardless of the type of smile.

However, the perception of smiles is not necessarily conscious. In [Rehm and André 2005], a perception test enabled the authors to measure the impact of polite smile expressions on participants subjective impressions of an ECA. The participants were able to perceive the difference, but they were unable to explain their judgment. These results are in line with the recent work of [Krämer et al. 2013] showing that users do

not consciously notice ECA smiles even if the virtual smiles have a significant effect on the users.

Finally, research shows that smiles expressed by both humans or virtual agents enhance the social stances perceived by others, particularly for smiling males (virtual or human) and for amused smiles. However, existing research has mainly compared the global perception of an agent expressing no smile, or, either an amused or a polite smile. One important difference in our work is that we investigate the effect of a virtual agent displaying both smiles types at different moments of the interaction. We aim at endowing an ECA with a kind of "smile focused theory of mind" to automatically infer, during the interaction, the potential user's perception of its social stances depending on its smiling behavior (defined by its smile type, when it smiles, in response to what). Indeed, an important factor in understanding the reaction of a user to a smile within the human context involves having a model of how a person may be expected to react to a specific smile type in a given circumstance. Any deviation from these norms of behavior would mean that something was probably not running smoothly with the conversation and some appropriate repair strategy or a change of conversation topic may be needed. In effect, this comparison of the behavior of an interlocutor against a model of socially normative behavior is part of a perspective taking mechanism or a smile focused theory of mind—the interlocutors are engaged in a mind-reading exercise in which deviations from normative behavior represent clues to changes in interpersonal stances. Such perspective taking and mind-reading behaviors have been recognized as important factors in human communication for some time [Whiten et al. 1997; Tomasello 1999; Dunbar 2001; McKeown 2013]. In this article, we present a first attempt to develop such a *smile focused theory of mind* model (Section 5).

The next section concentrates on the specification and integration of the proposed computational model of smiles into an ECA; the model includes components for the generation of smiles and smile behavior and a model that estimates a smile's effect from the perspective of the user—a rudimentary theory of mind targeted at the specific social signal of smiles.

3. FROM EMBODIED CONVERSATIONAL AGENT'S SMILES TO INTERPERSONAL STANCES

Our objective is to develop a computational model to enable an ECA to select and display the appropriate smiles during an interaction with a user. To develop such a model, smile expressions need to be considered from both production and perception perspectives:

(1) from a *production perspective*: the smile signal itself conveys different meanings depending on its morphological and dynamic characteristics (Section 2.1). A smile expression can enable an ECA to convey or reinforce a particular message to the user. For example, an amusement smile can communicate information to the user about the emotional state of the virtual agent. Consequently, the ECA should select and display the smile signal appropriate to its communicative intention. If the virtual agent has the intention to show embarrassment, the virtual agent's face should be animated accordingly to maximize the likelihood that it is perceived by the user as embarrassed.

(2) from a *perception perspective*: during the interaction, the expressed smile type combined with the situation in which the smile is expressed—the context—influences user perception of the ECA's interpersonal stance (Section 2.2.2). The virtual agent should then select its smiling behavior (in particular when to express a given smile) depending on the stances it wants to convey to the user. This adds a component of

interactional complexity by including contextual factors beyond the smile social signal; it acknowledges that a smile takes place within a broader social and perceptual context.

One aim of the work presented in this paper was to embed a computational model of smiles in the Greta system [Bevacqua et al. 2010b]. Greta is an open source interactive system that enables one to develop 3D ECAs capable of communicating expressive verbal and nonverbal behaviors. Moreover, Greta has been integrated in the platform SEMAINE [Schröder 2010]. This open-source platform enables users to naturally converse with ECAs. The SEMAINE platform adds modules to Greta that provide for the management of a real-time dialog with users. The SEMAINE platform was developed specifically for the purpose of enabling artificial listener interactions [Schröder et al. 2012], like the well-known Eliza chatbot [Weizenbaum 1966], with rich non-verbal behavior but poor linguistic competences. In SEMAINE, the communicative intention of the virtual character is automatically selected based on the analysis of user's visual and acoustic behaviors (for more details, see [Schröder 2010]). The non-verbal behavior of the virtual characters are determined by the Greta system. The architecture of Greta is based on SAIBA [Kopp et al. 2006], an international common multimodal behavior generation framework. The architecture is illustrated Figure 2. The *intent planner* generates the communicative intentions
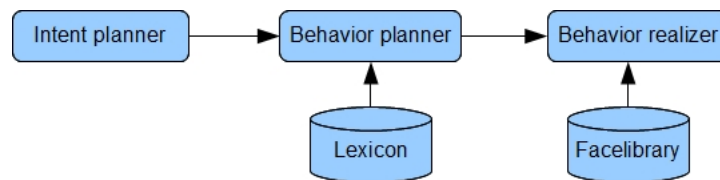

Fig. 2.   SAIBA architecture

(what the agent intends to communicate). For instance, a communicative intention can be the expression of joy. The *behavior planner* transforms these communicative intentions into a set of multimodal signals (e.g. speech, gestures, facial expressions). It is also responsible for synchronizing the signals across the modalities. The appropriate signals that can be used to express a particular intention are defined in the *lexicon*. The morphological and dynamic characteristics to adjust the face for the appropriate signals are described in the *facelibrary*. Finally the *behavior realizer* outputs the animation parameters for each of these signals.

Our objective was to give the capability to ECAs (such as those developed in the Greta system) to automatically determine their smiling behavior depending on the interpersonal stance they wants to convey to the user. For this purpose, the virtual agents have to have the capability to determine:

(1) how to display different smile types?
(2) when and which type of smile should be expressed during an interaction?
(3) what impact will the expression or non-expression of a smile have on user perception?

In order to respond to the first problem—*how to display different smile types*—the *Facelibrary* needs to be enriched with the different types of smile. In the proposed model, we have focused on three smile types: *amused*, *polite* and *embarrassed* (Section 2.1). In the following section, we present a user-perception approach to identify the

morphological and dynamic characteristics of these smile types. The smile characteristics identified using this approach were then integrated into the *Facelibrary* giving the ECAs the ability to express different smile types.

Addressing the second problem—*when and which type of smile should be expressed during an interaction*—requires linking the meaning conveyed by each smile type to the ECA's communicative intentions (the *which* question). The amused smile is linked to the communication of affective-like intentions such as happiness and liking; the polite smile is associated with the communication of cognitive-like intentions such as greetings, agreement, praise, encouragement, acceptance or understanding[4]. An embarrassed smile may be displayed to communicate embarrassment. Concerning the *when* question, an ECA may decide that it is not appropriate to express a smile associated to a communicative intention. To determine if a smile is expressed, a third issue needs to be addressed.

This third issue—*what impact will the expression or non-expression of a smile have on user perception*—required adding a module to the architecture to automatically infer the effect of a smile expression or lack of smile expression on user perception. Based on this module, the *behavior planner* may select the appropriate smiling behavior given an interpersonal stance to convey. This module and the methodology used to developed the smiling behavior, is described in detail in Section 5.

4. EMBODIED CONVERSATIONAL AGENT SMILES

As humans have the capability to express a variety of smiles during interpersonal interactions [Ekman 2009], smiling ECAs should ideally be endowed with some variety in smiling behavior. As highlighted in Section 2.1 smile signals can be categorized depending on the meaning they convey; we focus on three smile types from the range of human smiles: *amused*, *polite*, and *embarrassed* smiles. A first necessary step is to find out *how to display different smile types*, that is we need to identify how the ECA's face should be animated to convey either an amused, embarrassed, or polite smile.

4.1. User-created corpus of virtual agent smiles

One of the major challenges in the recognition and synthesis of verbal or non-verbal behavior is collecting datasets. It can be difficult, time consuming and expensive to collect and annotate them. A growing interest in using *crowdsourcing* to collect and annotate datasets has been observed in recent years [Yuen et al. 2011]. *Crowdsourcing* consists of outsourcing tasks to an undefined distributed group of people, often using the internet to recruit participants informally or through formal paid mechanisms such as Amazon's Mechanical Turk [Mason and Suri 2011]. In order to identify the morphological and dynamic characteristics of different virtual agent's smiles, we have used an informal crowdsourcing technique. We developed a crowdsourcing website, named *E-smiles-creator*, to collect a corpus of virtual agent's smiles directly created and annotated by users. The requested tasks in the E-smiles-creator consist of creating different smiles on an ECA's face.

E-smiles creator. The interface of the crowdsourcing website to create virtual smiles, called the *E-smiles-creator*, is composed of 4 parts (Figure 3):

(1) the upper section contains a description of the task: the smile that the user has to create, for instance, an amused smile;
(2) the left section contains a video showing the virtual agent animation, in a loop;

---

[4]Expressions on other modalities, such as gaze or gesture, are also associated to communicative intentions [Poggi and Pelachaud 2000; Heylen 2006]

(3) the right section contains a panel with the different smile parameters (such as its duration) that the user can change to create the smile (the animation on the left changes accordingly);
(4) the bottom part contains a Likert scale that allows users to indicate their satisfaction with the created smile.
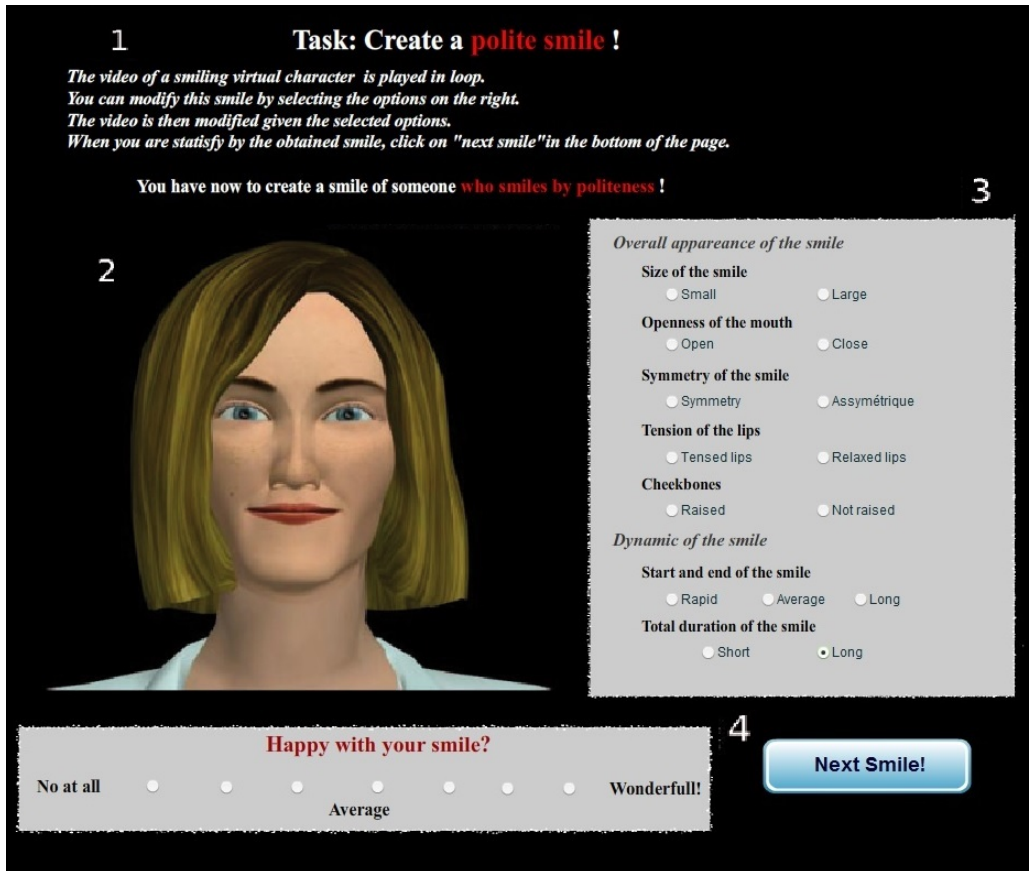
Fig. 3. Screenshot of the *E-smiles-creator* (originally the interface of the application was in French)

Using *E-smiles-creator*, a user can generate any smile by choosing the combination of seven parameters. Any time the values of one of the parameters changes, a corresponding animation is automatically played. Based on human smile research previously reviewed (see Section 2.1), we have considered the following morphological and dynamic characteristics of a smile: AU6 (cheek raising), AU24 (lip press), AU12 (zygomatic major), the symmetry of the lip corners, AU25 (mouth opening), the amplitude of the smile, the duration of the smile and the velocity of the rise and of the decay of the smile. Accordingly, on the right part of the *E-smiles-creator* interface (Figure 3, panel 3), the user may select values for these smile parameters. The animation of the smiling virtual face corresponds to a smile with the selected parameter values. We have

considered two or three discrete values[5] for each of these parameters: small or large smile (for the amplitude); open or close mouth; symmetric or asymmetric smile; tensed or relaxed lips (for the AU24); cheekbone raised or not raised (for the AU6); short (1.6 seconds) or long (3 seconds) total duration of the smile, and short (0.1 seconds), average (0.4 seconds) or long (0.8 seconds) beginning and ending of the smile (for the rise and decay)[6]. Considering all the possible combinations of these discrete values gives rise to 192 different animations of a smiling virtual agent face.

The *E-smiles-creator* was created with a French interface using Flash technology to enable broad distribution on the web. Users were asked to create one animation for each smile type; they then had to rate each created smile according to their level of satisfaction with the smile. The order of smiles and the initial values of the seven parameters were chosen randomly. To ensure the usability of the E-smiles-creator interface, we ran a pre-test with some participants (N=10). A short interview with each participant of the pre-test has enabled us to validate the interface.

This new method of creating facial expressions differs from existing approaches in several ways. Commonly, expressions are first created by researchers (possibly randomly) and then participants rate them (see Section 2.1). The current method involves the selection of a set of key features from the literature and allows participants to create the expression they believe corresponds to a given meaning. The combinations of parameters create a finite set of video animations, however, this still represents a considerable search space. Rather than forcing the users to view and rate each of the video clips in turn, the interface allows the users to intuitively constrain the search space and quickly narrow in on an appropriate animation.

Participants. Using the web interface, 348 mainly French people created smiles (195 females; mean age of 30 years). Each participant has created one smile for each category (amused, polite and embarrassed). We then collected 1044 smile descriptions: 348 descriptions for each smile (amused, polite and embarrassed).

Results. On average, participants were satisfied with the created smiles (5.28 on a Likert scale of 7 points). User satisfaction was similar for the three smile types (between 5.2 and 5.5). Collating the values users selected for the 7 parameters when creating each of the 3 smiles, we can derive common characteristics for each of them. Amused smiles are mainly characterized by large amplitude (83.6% of the total amused smiles), an open mouth (85.6%), and relaxed lips (92.2%). Most amused smiles also contained cheek raising (78.4%) and a long global duration (84.4 %). Compared to amused smiles, embarrassed smiles often have small amplitude (73.1% of the total embarrassed smiles), a closed mouth (81.8 %), and pressed lips (74.6 %). They are also characterized by the absence of cheek raising (59 %). Polite smiles were mainly characterized by small amplitude (67.7 % of the total polite smiles), a closed mouth (76 %), symmetry (67.1 %), relaxed lips (69.4 %), and an absence of cheek raising (58.9 %)(for more details on the corpus of smiles, see [Ochs et al. 2010]).

---

[5]The selection of discrete values through radio buttons, rather than continuous values through sliders, has been chosen to minimize the complexity of the task (both from the user point of view and from a technological point of view since each smiling facial animation has to be created before and stored in the application). However, to obtain a more fine-grained description of smiles, continuous variables could be considered.
[6]The values of the rise and the decay have been defined to be consistent with the values of the duration of the smile.

Based on this smile corpus and on a decision tree classification technique, described in the next section, we have extracted the morphological and dynamic characteristics of the smile types that a virtual agent may express.

4.2. The morphological and dynamic characteristics of smiles

The *E-smiles-creator* smile corpus provided a large body of data that could be applied to selecting appropriate facial expressions for an ECA to display in order to express amused, polite, or embarrassed smiles. For this purpose, we adopted a machine learning technique to determine the most suitable morphological and dynamic characteristics of these smiles.

Decision tree. Amongst the different machine learning techniques, we used a *decision tree learning algorithm* to identify the different characteristics of the amused, polite, and embarrassed smiles in the corpus. This choice is motivated by the fact that the decision tree is a "white box" model in the sense that the acquired knowledge is expressed in a readable form [Su et al. 2008]. Indeed, the produced results in the form of a tree are easily understood and interpretable. The decision tree can be translated into a set of rules by creating a separate rule for each path from the root to the leaf. In the ECA architecture (Section 3), the morphological and dynamic characteristics of the signals are explicitly represented in the *Facelibrary*. Other machine learning techniques based on "black box" models, such as Support Vector Machines (SVM) or Neural Networks, are not as well-adapted to our research goal since the results of such algorithms are difficult to understand and to represent explicitly.

In the decision tree algorithm, the input variables (predictive variables) are the morphological and dynamic characteristics and the target variables are the smile types (amused, polite, or embarrassed). Consequently, the nodes of the decision tree correspond to the smile characteristics and the leaves are the smile types.

The user indicated *level of satisfaction* for each created smile (a level that varied between 1 and 7). This level of satisfaction was used to create the decision tree, with the assumption that smiles with a high level of satisfaction were more reliable than low level smiles. In order to create a range of non-equivalent smiles in the corpus, we *oversampled* to give a higher weight to the smiles with a high level of satisfaction[7]: each created smile was duplicated n times, where n is the level of satisfaction associated with this smile. So, a smile with a level of satisfaction of 7 is duplicated 7 times whereas a smile with a level of satisfaction of 1 is not duplicated. The resulting data set is composed of 5517 descriptions of smiles: 2057 amused smiles, 1675 polite smiles, and 1785 embarrassed smiles.

Results. The free data mining software TANAGRA [Rakotomalala 2005] was used to construct the decision tree; using a CART (Classification And Regression Tree) approach [Breiman et al. 1984]. The resulting decision tree is represented in Figure 4. The tree is composed of 39 nodes and 20 leaves. All the input variables (the smile characteristics) were used in the classification. The structure of the tree informs us about the *discriminative power* of the variables. Among the 7 variables, the openness of the mouth, the smile size, and the lips tension, at the top of the tree, correspond to the most important variables in distinguishing amused, polite and

---

[7]Note that an undersampling could be also applied to the corpus by not considering smiles with a low level of satisfaction. We have given priority to oversampling to have the maximum of smile descriptions to model the variability of smile expressions.

Decision tree

- **MOUTH_OPEN in [open]**
  - ○ SMILE_SIZE in [small]
    - ■ LIPS_TENSION in [no_tension]
      - ■ SMILE_DURATION in [short] then SMILE_TYPE = **polite** (60,40 % of 101 examples) ◀
      - ■ SMILE_DURATION in [long]
        - ■ SMILE_SYM in [sym]
          - ■ SMILE_AU6 in [no_AU6] then SMILE_TYPE = **polite** (66,00 % of 50 examples)
          - ■ SMILE_AU6 in [AU6] then SMILE_TYPE = **amused** (64,58 % of 48 examples)
        - ■ SMILE_SYM in [asym] then SMILE_TYPE = **amused** (60,53 % of 76 examples)
    - ■ LIPS_TENSION in [tension]
      - ■ SMILE_SYM in [sym] then SMILE_TYPE = **polite** (61,11 % of 54 examples)
      - ■ SMILE_SYM in [asym] then SMILE_TYPE = **embarrassed** (69,09 % of 55 examples)
  - ○ SMILE_SIZE in [large]
    - ■ LIPS_TENSION in [no_tension]
      - ■ SMILE_DURATION in [short]
        - ■ SMILE_AU6 in [no_AU6] then SMILE_TYPE = **polite** (52,63 % of 57 examples)
        - ■ SMILE_AU6 in [AU6] then SMILE_TYPE = **amused** (73,08 % of 130 examples)
      - ■ SMILE_DURATION in [long] then SMILE_TYPE = **amused** (89,38 % of 951 examples)
    - ■ LIPS_TENSION in [tension]
      - ■ SMILE_ONOFFSET in [short,medium] then SMILE_TYPE = **amused** (41,67 % of 34 examples)
      - ■ SMILE_ONOFFSET in [long] then SMILE_TYPE = **embarrassed** (61,29 % of 31 examples)
- **MOUTH_OPEN in [close]**
  - ○ LIPS_TENSION in [no_tension]
    - ■ SMILE_SYM in [sym]
      - ■ SMILE_AU6 in [no_AU6] then SMILE_TYPE = **polite** (84,05 % of 370 examples)
      - ■ SMILE_AU6 in [AU6]
        - ■ SMILE_SIZE in [small] then SMILE_TYPE = **polite** (66,12 % of 121 examples)
        - ■ SMILE_SIZE in [large]
          - ■ SMILE_ONOFFSET in [short,medium] then SMILE_TYPE = **amused** (45,00 % of 36 examples)
          - ■ SMILE_ONOFFSET in [long] then SMILE_TYPE = **polite** (67,27 % of 55 examples)
    - ■ SMILE_SYM in [asym]
      - ■ SMILE_DURATION in [short] then SMILE_TYPE = **polite** (61,84 % of 152 examples)
      - ■ SMILE_DURATION in [long]
        - ■ SMILE_SIZE in [small] then SMILE_TYPE = **polite** (53,72 % of 121 examples)
        - ■ SMILE_SIZE in [large]
          - ■ SMILE_AU6 in [no_AU6] then SMILE_TYPE = **amused** (57,41 % of 54 examples)
          - ■ SMILE_AU6 in [AU6] then SMILE_TYPE = **polite** (41,46 % of 41 examples)
  - ○ LIPS_TENSION in [tension] then SMILE_TYPE = **embarrassed** (70,60 % of 1085 examples)
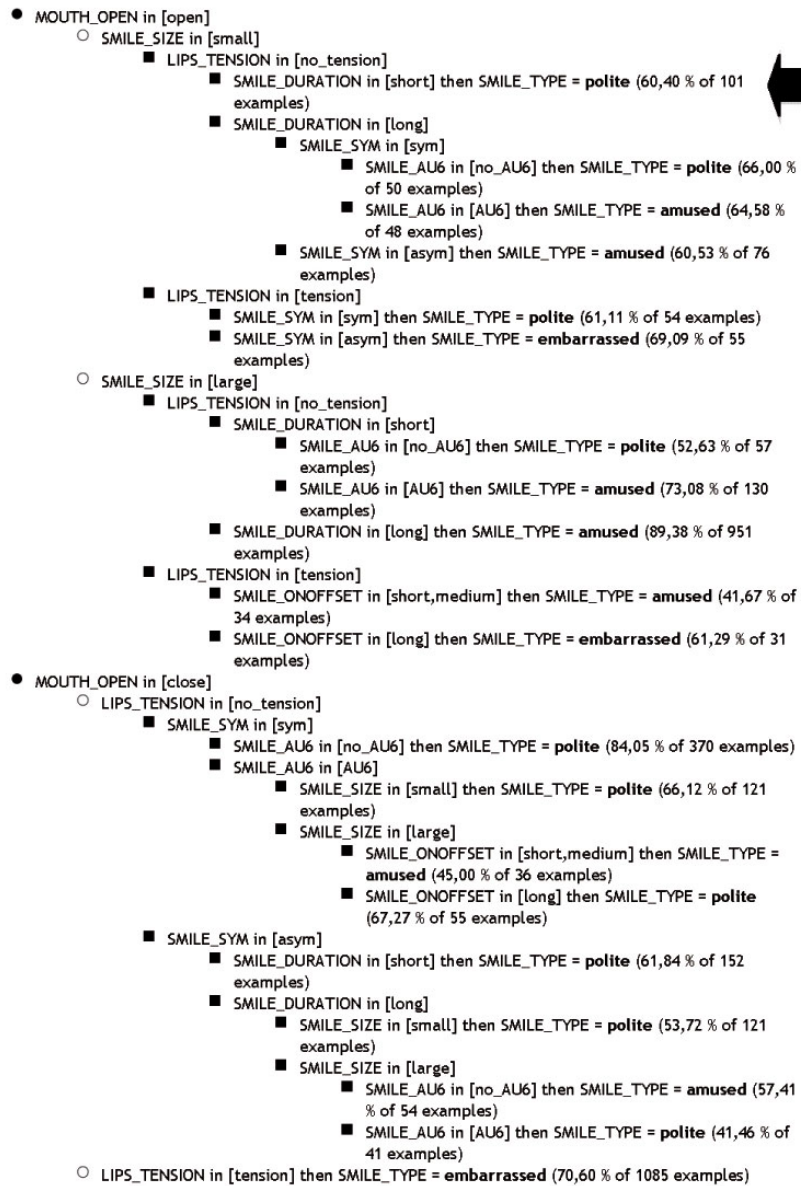
Fig. 4. Smiles Decision Tree. At each line, the term within brackets corresponds to the value of the characteristic: *open* or *close* for the openness of the mouth, *small* or *large* for the size of the smile, *no_tension* or *tension* for the lips tension, *short* or *long* for the duration of the smile, *sym* or *asym* for the symmetry of the smile, *no_AU6* or *AU6* for the cheek raising, and *short*, *medium* or *long* for the duration of the rise and decay. The terms in bold represent the smile type with the values of the parameters leading to this leaf in the tree. The numerical values within the parentheses correspond to the percentage of well-classified smiles in this category and the total smiles that fall within this category.

embarrassed smiles[8]. The values within parentheses at each leaf correspond to the percentage of well-classified smiles in this category and the total smiles that fall within this category. For example, for the leaf indicated by a black arrow Figure 4, 101 smiles have the following characteristics: an open mouth, a small size, lips pressed and a short duration; and 61% of these 101 smiles are classified as polite. The global error rate is 27.75%. With 95% confidence intervals at ±1.2%: the global error rate is in the interval [26.55%, 28.95%]. An analysis of the error rate for each smile type shows that the amused smiles are better classified (18% of error with a confidence interval of ±1.8%) than the polite (34% of error with a confidence interval of ±1.7%) and the embarrassed smiles (31% of error with a confidence interval of ±1.7%).

4.3. The lexicon of embodied conversational agent's smiles

The smiles decision tree revealed 20 different *smile patterns*, corresponding to the 20 leaves of the tree (Figure 4). Ten leaves are labeled as polite smiles, 7 as amused smiles, and 3 as embarrassed smiles. Because some branches of the tree do not contain a value for each morphological and dynamic characteristic, more than 20 smiles could be created from our decision tree. For instance, for the first polite smile pattern that appears in the tree (indicated by a black arrow on Figure 4), the size of the smile, its duration, and the velocity of the rise and decay are not specified. Consequently, this polite smile pattern can be expressed by the virtual agent in 12 different manners.

In order to identify the smile that the ECA should express, in the decision tree, we have selected the leaves leading to the best classified amused and polite smiles (4 leaves for each smile) and the leaves leading to the best classified embarrassed smiles (3 leaves). The objective is to consider, not only one amused, polite, or embarrassed smile but different possible expressions for each smile type. That enables us to increase the variability of the virtual agent's expressions: the virtual agent may express the same type of smile in different manners during an interaction to avoid repetition of the exact smile pattern. Previous research has shown that the non-repetitive behaviors of an ECA improves the perceived believability [Niewiadomski et al. 2011].

To create the lexicon of ECA's smiles, for each leaf of the tree, we compute the 95% confidence interval from the percentage of well-classified smiles of the leaf and the number of smiles that fall within the leaf (Figure 4) using the formula:

$$r = 1.96\sqrt{\frac{p * (1 - p)}{N}} * 100$$

such as N is the number of smiles that fall within the leaf and p the percentage of well-classified smiles. The 95% confidence interval is then $[p - r, p + r]$. For instance, for the first polite smile appearing in the tree (indicated by a black arrow on Figure 4), the percentage of well-classified smiles is 60.41% and the number of smiles that fall within the leaf is 101 (Figure 4). The 95% confidence interval for this leaf is $[60.41 - 9.5, 60.41 + 9.5] = [50, 91; 69, 91]$. The confidence interval enables us to consider the number of well and badly-classified smiles instead of considering only the probability. For the amused and polite smiles, we have then selected the four smile patterns that have the confidence intervals with the highest suprenum (upper bound) and with the smallest size. For the embarrassed smiles, we have selected the three smile patterns of embarrassment. To determine the smile's characteristics of these smile patterns that are not defined in the tree, we consider the *contingency table* representing the

---

[8]One disadvantage of the decision tree method is the robustness: a change in the input data may change the tree structure. To overcome this limitation, we have conducted perceptive tests to ensure that the identified smiles with their characteristics are correctly recognized by the users (Section 4.4).

frequency of smile types for each characteristic (Table I). For example, 16.4% of the

Table I. Contingency table of the smile's characteristics and the smile types

variable	value	amused	embarrassed	polite
size	small	16,4%	73,1%	67,7%
	big	83,6%	26,9%	32,3%
mouth	close	14.4%	81.8%	76%
	open	85,6%	18,2%	24%
symmetry	sym.	59,9%	40,5%	67,1%
	assym.	40,4%	59,1%	32,9%
lips tension	no tension	92.2%	25.4%	69.4%
	tension	7.8%	74.6%	30.6%
AU6	no	21.6%	59%	58.9%
	yes	78.4%	41%	41.1%
rise/decay	short	33.4%	28.9%	30.3%
	average	30.3%	39.6%	37.1%
	long	36.3%	31,5%	32.6%
duration	short	15.6%	43.6%	42.9%
	long	84.4%	56.4%	57.1%

total amused smiles, 73.1% of the total embarrassed smiles, and 67.7% of the total polite smiles have a small size. For instance, if the selected smile pattern is the first polite smile that appears in the tree (indicated by a black arrow on Figure 4), the following characteristics are not specified in the tree: the size of the smile, its duration, and the velocity of the rise and decay. To fill in these values, we refer to the contingency table. Because in this table, it appears that a majority of polite smiles have a small size, long duration, and an average velocity of the rise and decay, we consider a smile with such characteristics and the characteristics described in the branch of the tree leading to the selected smile. The characteristics of the selected smiles following the algorithm described above are illustrated Table II. For embarrassment, only three different smile patterns exist in the tree. To balance the number of smiles for each type, the fourth embarrassed smile of the Table II was generated using the smile pattern of the smile with the id. 11 but three of its characteristics were chosen to be opposite to the ones indicated by the contingency table.

Table II. The characteristics of the selected smiles

id	type	size	mouth	symmetry	lip press	cheek raising	rise decay	duration
1	polite	small	close	yes	no	no	0.4s	3s
2	polite	large	close	yes	no	yes	0.8s	3s
3	polite	small	close	yes	no	yes	0.4s	3s
4	polite	small	close	no	no	no	0.4s	1.6s
5	amused	large	open	yes	no	yes	0.8s	3s
6	amused	large	open	yes	no	yes	0.8s	1.6s
7	amused	small	open	yes	no	yes	0.8s	3s
8	amused	small	open	no	no	yes	0.8s	3s
9	embarass.	small	open	no	yes	no	0.4s	3s
10	embarass.	small	close	no	yes	no	0.4s	3s
11	embarass.	small	close	no	yes	no	0.1s	3s
12	embarass.	large	open	no	yes	no	0.8s	3s

The decision tree approach and the proposed algorithm permit a better representation of user perception in comparison to methods that select the most frequently created smile types. For instance, if we look at the 4 most frequently created embarrassed smiles, these only take into consideration the perception of 66 participants who created through the E-smiles-creator interface. On the other hand, using a decision

tree approach allows us to encompass all participants' created smiles. This approach enables us to not only treat each created smile as a whole but to also provide insight concerning the contribution of each feature to the created smiles. This more inclusive approach considers not only the perception of particular users but the perception of all the users who have created smiles. Moreover, the decision tree approach allows better variability in smile expressions. For instance, the four most frequent amused smiles differ only in two characteristics (symmetry and rise/decay) whereas the four amused smiles resulting from our algorithm differ in three characteristics (size, symmetry and duration).

We have now enhanced the lexicon of the ECA with a variety of smiles. However these smiles have been created without considering context. The following section presents a perception study to ensure that the selected smiles are appropriate in amused, polite or embarrassing situations.

4.4. Perceptual validation of the virtual agent smiles in context

A perceptual evaluation of the selected amused, polite, and embarrassed smiles was performed to validate the smiles (Table II). A similar informal crowdsourcing approach (see Section 4.1) was used through, once again, an interface developed using the Flash technology platform. Different scenarios of polite, amused, and embarrassed situations were presented in written text to the user. Each scenario presented a woman named *Greta* involved in a particular situation and activity. The scenarios were 3-4 sentences long. An example of the politeness scenarios is "Greta was a hosting a party at her home. The guest list included many people who didn't know one another. As her guests began to arrive, Greta would greet them at the door and take their coats. Before moving on to other guests, she would offer them a drink and introduced them to other guests present". To ensure that the scenarios reliably represented the intended states, they were tested with a panel of participants (for more details see [Ochs et al. 2012]).

In the test, six scenarios (from a total of twelve) were presented to the user. For each scenario, three video clips of different virtual agent smiles were displayed. Users were asked to imagine the virtual agent displaying the facial expression in the situation presented in the scenarios. Then they had to rate each of the three facial expressions on its appropriateness for the given scenario and to rank them in order of appropriateness (Figure 5).

Seventy-five individuals participated in this evaluation (57 females, mean age of 32). The mean ratings of appropriateness (and standard errors) of each smile in the politeness, amusement, and embarrassment scenarios respectively are presented in Figure 6. The evaluation revealed significant differences showing that most of the generated smiles are appropriate to their corresponding context. Only for 2 out of the 12 smiles did the results indicate that the smiles may not match their intended context. One of them, smile 12, corresponds to the last embarrassed smile pattern—that has a relatively high error rate. Smile 2 was rated higher in amusement scenarios than in the politeness scenarios. This lead us to conclude that the large intensity that characterizes smile 2 compared to the other polite smiles, resulted in the participants interpreting it differently than the other polite smiles (for more details on the experiment, see [Ochs et al. 2012]).

Comparisons of validated smiles with related works. The generated and validated smiles (Table II) are fairly consistent with the studies on human smiles (Section 2.1). The lip press occurs only in embarrassment smiles which is consistent with Keltner's findings [Keltner 1995]. Amused smiles are characterized by cheek raising as stated in [Ekman 2003]. Cheek raising may also appear in the expression of polite smile as claimed in [Krumhuber and Manstead 2009]. Mouth opening occurs in all amusement
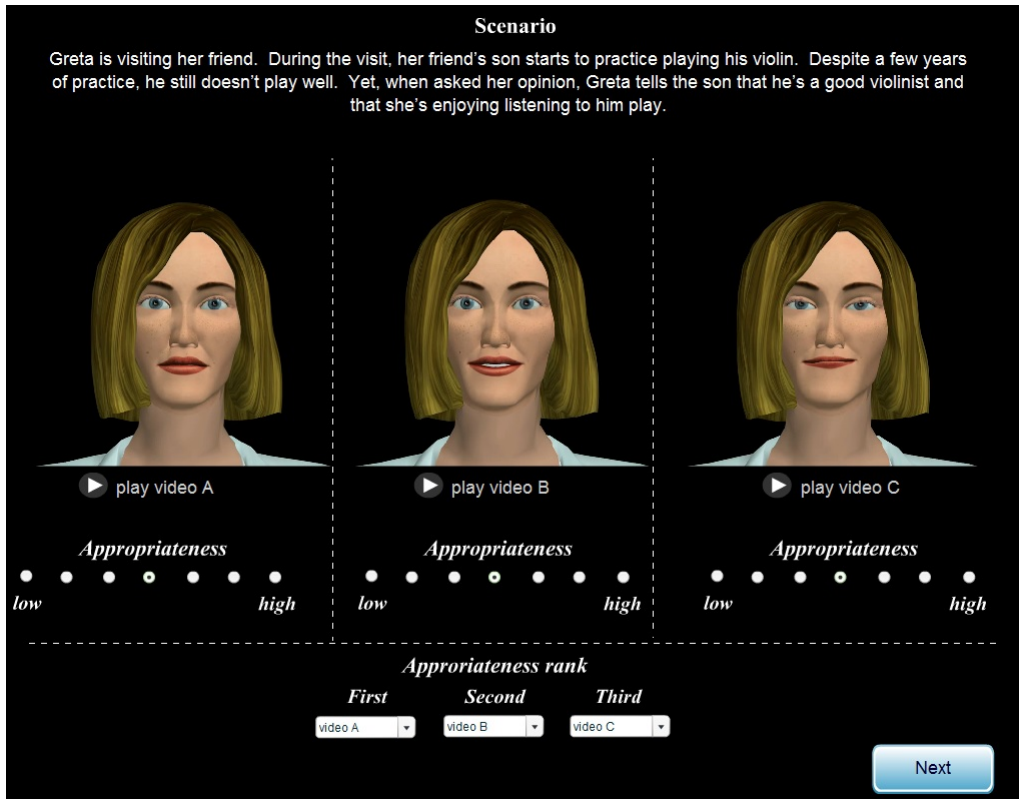
Fig. 5.   Screenshot of the interface

id	Politeness	Amusement	Embarrassment
1	4.74 (1.45)	3.75 (1.51)	3.93 (1.64)
2	3.35 (1.75)	4.41 (1.82)	3.49 (1.83)
3	5.06 (1.49)	3.43 (1.52)	4.81 (1.50)
4	4.48 (1.79)	2.83 (1.75)	3.74 (1.84)
5	3.56 (1.80)	6.01 (1.12)	3.70 (1.98)
6	3.14 (1.75)	4.73 (1.68)	4.15 (1.73)
7	5.23 (1.48)	4.93 (1.62)	2.95 (1.65)
8	2.75 (1.42)	4.53 (1.61)	4.04 (1.78)
9	3.30 (1.43)	4.23 (1.78)	4.29 (1.98)
10	5.26 (1.82)	2.06 (1.51)	4.91 (1.60)
11	4.56 (1.93)	2.28 (1.59)	4.71 (1.84)
12	3.91 (1.89)	5.15 (1.65)	4.24 (1.75)

Fig. 6.   Mean (and standard deviations in parentheses) ratings of appropriateness of each smile in politeness scenarios, amusement scenarios, and embarrassment scenarios. Smile numbering matches labels presented in Table II

smiles, but it also occurs more often in embarrassed smiles than in polite smiles [Ambadar et al. 2009]. Amusement smiles are also more often symmetric (3 out of 4 cases) than other smiles [Ekman 2003]. Amusement smiles often have longer rise and decay than other smiles which is consistent with findings by [Hoque et al. 2011; Ekman and Friesen 1982]. This suggests that users in this study and the previous techniques in the literature are drawing upon a common pool of knowledge and that, in this study, common knowledge of human smiles is being applied to virtual agents.

The 10 validated smiles have been integrated in the *Facelibrary* of the ECA architecture (presented Section 3) to enable the ECAs to display these smiles during an interaction. As described in Section 3, the smiles are displayed by the ECA in accordance with its communicative intentions. For instance, an amused smile is expressed when the virtual agent has the intention to communicate a positive emotion of happiness. However, there are situations when withholding an expression may carry communicative weight in conveying a particular interpersonal stance. To give such a capability to an ECA, we have constructed a model to automatically compute the potential perception of the user based on the smiling behavior of the virtual agent. The next section presents this model in more details.

5. EMBODIED CONVERSATIONAL AGENT'S EXPRESSION OF INTERPERSONAL STANCES THROUGH SMILES

The expression of smiles is not essential to communicate an intention. The verbal message, with a neutral facial expression, may be sufficient to deliver the information. For instance, an ECA may decide to express joy without the display of a smile but only with an utterance such as "I'm happy today!", although such an utterance would be highly likely to co-occur with a smile. In the same way, an expression of understanding or agreement may or may not be accompanied with a polite smile. As highlighted in Section 2.2, the display of smiles or the non-expression of a smile in situations in which a smile might be expected impacts an interlocutor's perception of the other interlocutor's interpersonal stance. The smiling behavior of the virtual agent during an interaction—the expression or non-expression of a smile—should be selected depending on the interpersonal stance that the virtual agent wants to express. Selecting the appropriate smiling behavior therefore requires the ECA to estimate the effects of its smiling behavior on user perception, i.e. to have a kind of "smile focused theory of mind" model (Section 2.2). We have constructed such a model that automatically computes an estimation of user perception of an agent's interpersonal stances during the interaction. The development of this model involved creating a probabilistic model of user perception of the interpersonal stances occurring in interactions with smiling ECAs. There are two stages to the development of this model:

(1) first of all, the effects of ECA's smiling behavior on the user perceptions of its interpersonal stances should be identified. Indeed, research on this topic (Section 2.2) has mainly focused on the comparison of the effects of a smiling and non-smiling virtual agent on the global user's perception of the agent. The previous studies did not consider smiling behavior combining different types of smile. Consequently, we have collected users' perceptions of ECAs displaying different smiling behaviors (Section 5.1.1);
(2) based on the collected users' perceptions, we have implemented a probabilistic model to estimate the user's perception of the ECA's interpersonnal stances in real-time during the interaction (Section 5.1.2).

The model was then experimentally validated (Section 5.2).

5.1. A probabilistic model of user perceptions of smiling ECA interpersonal stances

5.1.1. User perceptions of smiling ECA interpersonal stances. User perceptions of virtual agents displaying polite and amused smiles were examined to measure the effects of the agents' smile expressions. We focus on situations in which smiles are expressed when an ECA is speaking[9], and on positive situations: using amused and polite smiles. The amused smile situation is 'telling a riddle' situation and the polite smile display accompanies the ECA's salutation at the beginning of its talk; this corresponds to greeting the user (as proposed in [Poggi and Chirico 1998; Cassell et al. 2001]). To balance the effect of the appearance, we use two different ECAs in these studies: one ECA with a female appearance, named *Poppy*, and one ECA with a male appearance, named *Obadiah* (Figure 7).

Fig. 7.   Screenshot of the two smiling embodied conversational agents

Procedure. User perception data was collected with a similar crowdsourcing method and Flash-based web platform to that discussed in Section 4.1. Initially, each participant watches four videos of the ECA telling a riddle (Figure 8): two video clips of the virtual agent *Poppy* and two video clips of the virtual agent *Obadiah* (Figure 7). The four riddles told to the participant were different. The order of the video clips was counterbalanced to avoid order effects on the results. The video clips presented to the participants involved a brief salutation followed by the virtual agents telling the user a riddle in French. For instance: "Good morning, I know a little riddle, what is the future of I yawn? I sleep!" (translated from French). Four different riddles have been selected based on a brief evaluation of sixteen riddles. We asked 7 individuals (3 females and 4 males) to rate their liking of the sixteen riddles between 0 and 5. Based on the results, we selected the riddles with the maximum rate and the minimum standard deviation. We assumed that the four selected riddles are approximately equivalent. Finally, in terms of verbal behavior of the virtual agent, only the riddle varies from one video clip to another. The beginning of the talk and the tonality of the voice do not vary. Concerning the non-verbal behavior, only the smiles (both the type of smile and the moment when they are expressed) differ from one video clip to another. There were four conditions:

— *no smile condition*: the ECA expresses no smile during its talk;
— *polite smile condition*: the ECA displays only the polite smile when the ECA is saying "good morning";

---

[9]We do not explore the display of smile when the virtual agent is listening, *i.e.* smiles used as backchannel. For instance see [Bevacqua et al. 2010a] for a study on its effect on user perception.
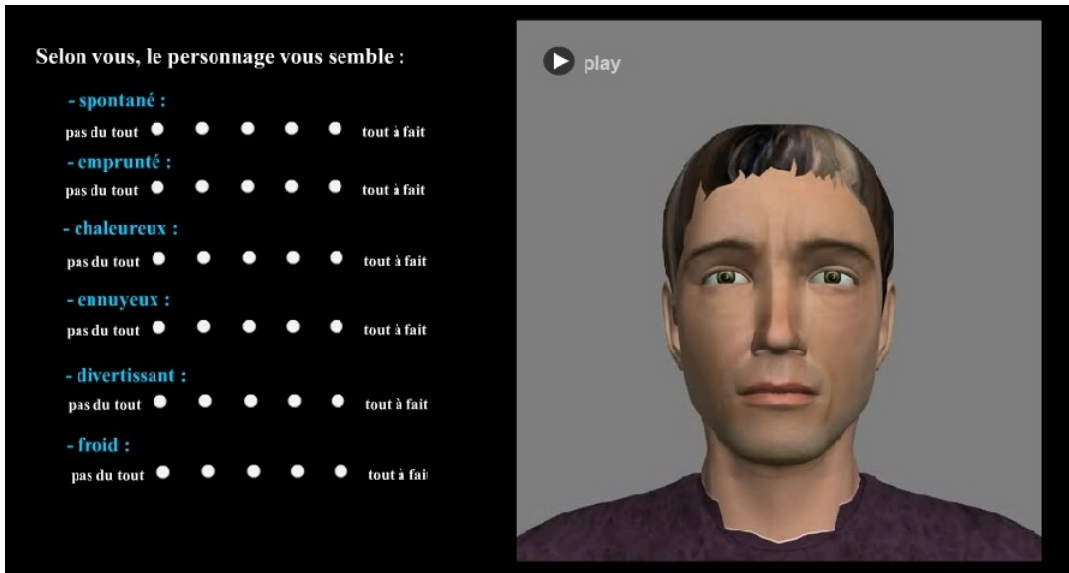
Fig. 8.   Screenshot of the interface to collect users' perception of virtual agents' social stances

— *amused smile condition*: the ECA expresses only the amused smile when it says the response to the riddle;
— *both smiles condition*: the ECA displays the polite and amused smiles at the moment described in the polite and amused conditions.

Note that the *type of smile* variable and *the moment when the smile is expressed* variable are not considered separately. The smile types are associated with particular communicative intentions expressed in specific moment in the video clips; we are not interested in this study in the effects of placing the alternative smile with an inappropriate communicative intention. As a result we have not considered the polite smile in the place appropriate for an amused communicative intention or conversely the amused smile in the place appropriate for a polite communicative intention.

After watching each video clip, participants had to rate the interpersonal stance of the ECA on a 5-point Likert scale. In this study, we considered the following stances as being relevant to the scenarios: *spontaneous*, *stiff*, *cold*, *warm*, *boring*, and *enjoyable*.

Only the virtual agent *Poppy*, a female character, was considered in the previous studies (Section 4). To verify that the smiles of the virtual agent *Obadiah*, a male character, are perceived by the users as expected (amused, polite and embarrassed as for the Poppy agent), in a second part of the test, four videos of the smiling *Obadiah* ECA were presented to the user. Here, the virtual agent just smiles without speaking. For each video, we asked the user to indicate the type of smile displayed by the virtual agent: *polite*, *amused*, *none of them*. Once again, the order of the presented videos was counterbalanced to avoid order effects.

Participants. Two hundred and forty two individuals participated in this study (158 female) with a mean age of 30 (SD = 10.35). They were recruited via online mailing lists. The participants were mainly from France (N = 223), followed by Belgium (N = 5). There were participants from Germany, Algeria, Tunisia, and Italy. Each

participant watched four video clips (two of Poppy and two of Obadiah each telling a different riddle) and four video clips of the virtual agents just smiling (in the second part of the study).

Results. The smile categorization task showed that the smiles were categorized correctly on average, except in one case—the amused smile displayed by the virtual agent *Poppy*—which was categorized more often as polite than amused (smile with id 6 in Table II)[10]. As a result the user ratings on the video clips in which *Poppy* displays this smile were excluded. In total 483 video clip ratings were considered.

The effects of smiles on user perception were analysed using ANOVAs and post hoc Tukey tests to assess differences in ratings between the conditions (no smile, polite smile, amused smile, and both smiles condition). Table III presents the results. The first two columns indicate the conditions being compared. The condition in the table cells refers to which of these two conditions received the higher rating for the given interpersonal stance (n.s. means non significant, *: $p < .05$, **: $p < .01$, ***: $p < .001$). For example, the first cell in the "warm" interpersonal stance column indicates Amused was perceived as highly significantly more "warm" than the No smile condition.

Table III. Comparison of the users' perception of the virtual agents' social stance in the different conditions

Conditon 1	Conditon 2	Warm	Enjoyable	Cold	Boring
No smile	Amused	Amused***	Amused*	No smile*	n.s.
No smile	Polite	Polite***	n.s.	n.s.	n.s.
No smile	Both	Both***	Both**	No smile***	No smile**
Polite	Both	Both***	Both*	n.s.	n.s.

The differences of user perceptions due to the appearance of the ECA—that is, whether it is Poppy or Obadiah— were assessed with t-tests. The results show that the appearance of the ECA has significant effects on the user's perception. For instance, when Poppy is smiling (either smile type), she is perceived as significantly less *cold* (and warmer) than Obadiah expressing the same smile ($p < 0.05$). Poppy is perceived as less *boring* when displaying one smile (either polite or amused) than Obadiah with the same smile ($p < 0.05$). With the amused smile (with or without a polite smile), Poppy is perceived significantly more *spontaneous* and *enjoyable* than Obadiah expressing the same smile ($p < 0.01$). With regard to these results, we have more precisely analyzed the significant differences for each virtual agent separately. Contrary to the results presented in Table III, it appears that, compared to the expression of only the polite smile, Poppy is perceived significantly more *spontaneous*, *warm* (and less *cold*), and less *stiff* when she expresses an amused smile ($p < 0.05$). For Obadiah, the expression of an amused smile (with or without a polite smile) enhances the warm impression of the virtual agent ($p < 0.05$).

The results reveal the importance of considering both the types of smile and the appearance of the ECA to compute the user perception of an agent's interpersonal stance. Based on this data we have constructed a model to automatically compute user perceptions of an ECA's interpersonal stance. This model is presented in detail in the next section.

---

[10]This smile seems to have been categorized as a polite smile because of its short duration. The forced-choice question in this study (that obliged users to select between polite smile, amused smile or none of them) may explain this difference with the results of the previous study.

5.1.2. Corpus-based probabilistic model of user perception of smiling ECA interpersonal stances.
We have argued that ECAs need to model how a user perceives them, in particular,
how a user perceives their interpersonal stance. This section proposes a model to
automatically compute an estimate of user perception that is dependent on the smile
displayed by the virtual agent and on its appearance. The model aims to estimate the
probability that a virtual agent is perceived as *spontaneous*, *stiff*, *warm*, *enjoyable* and
boring.

The work in the previous section resulted in user perception data for each of these
interpersonal stances along a 5-point Likert scale. We can represent these as natu-
ral values ranging from 1 to 5. In a mathematical point of view, interpersonal stances
are *fuzzy concepts*, i.e. lacking a fixed and precise meaning [Dietz and Moruzzi 2010].
Fuzzy variables are generally used to discuss fuzzy concept. A fuzzy variable is a value
which may range in an interval defined by quantitative limits and which can be use-
fully described with imprecise categories (such as "high" or "low"). In our model, we
have defined three fuzzy variables: neutral (associated to the value 1)[11], low (associ-
ated to the values 2 and 3), and high (associated to the values 4 and 5). Since our
model cannot produce outputs (the estimated user's perceptions of ECA's stances) with
certainty, probabilities for each of the fuzzy variables are computed. The probabili-
ties of obtaining the variables for each interpersonal stance are computed based on
the results of the study (previous section). For instance, the probability that the vir-
tual agent *Poppy* is perceived as highly spontaneous when displaying an amused smile
and telling something positive is $P(spontaneous = high|(smile = Amused \vee smile = Both) \wedge appearance = Poppy) = 0.27$, that is, the probability that spontaneous=4 or
spontaneous=5 in the Amused condition (only an amused smile is expressed) or Both
condition (an amused and polite smile are expressed).
Finally, given its appearance and its smiling behavior (polite smile, amused smile,
both smiles, or no smile), the model provides a matrix reflecting the probability of the
user's (potential) perception of the virtual agent's interpersonal stance. This matrix
evolves as the agent communicates various intentions. For instance, the matrix
illustrated in Figure 9 reflects the potential interpersonal stance perceived by the
user for the Obadiah appearance virtual agent saying something positive without
expressing a smile (*i.e.* the virtual agent has not expressed an amused smile in a
situation in which such a smile could be expected), at a time t of the interaction but
also globally at the end of the interaction. The model enables us to measure the effects
of a smile but also the effect of *not* displaying a specific smile in a situation in which
the user may expect this non-verbal behavior.

The model has been integrated in the Greta architecture. The SEMAINE platform
(Section 3) allows a user to interact vocally in natural language with the ECAs—where
the ECAs mainly adopt the conversational role of a listener—the user role is to initiate
a small-talk conversation with minimal constraints[12]. Figure 10 illustrates the output
of the module (the matrix of probabilities) during an interaction of a user with the
virtual agent Poppy; the integration of the model with the SEMAINE platform and
interface can also be seen in the figure. Depending on the smiling behavior of the vir-
tual agent, the matrix of probabilities—representing the potential user perception of

---

[11]The neutral variable means that the stance has not been perceived. This variable has been associated to
the value 1 since on the Likert scale presented to the participants (Figure 8), this value is labeled "Not at
all" to indicate the absence of expression of the stance.
[12]Note that in the version of SEMAINE used in this study, the non-verbal behavior of the user was not
detected.

	neutral	low	high
spontaneous	0.44	**0.47**	0.09
stiff	0.14	**0.49**	0.37
warm	**0.54**	0.42	0.04
enjoyable	0.37	**0.55**	0.09
boring	0.06	0.29	**0.65**
cold	0.05	0.28	**0.67**

Fig. 9. Matrix of probabilities representing the user's (potential) perception of the virtual agent *Obadiah* who does not display an amused smile when telling a riddle. The values strictly superior to 0.45 are highlighted in bold.

the virtual agent's interpersonal stance—is automatically updated. At each interac-
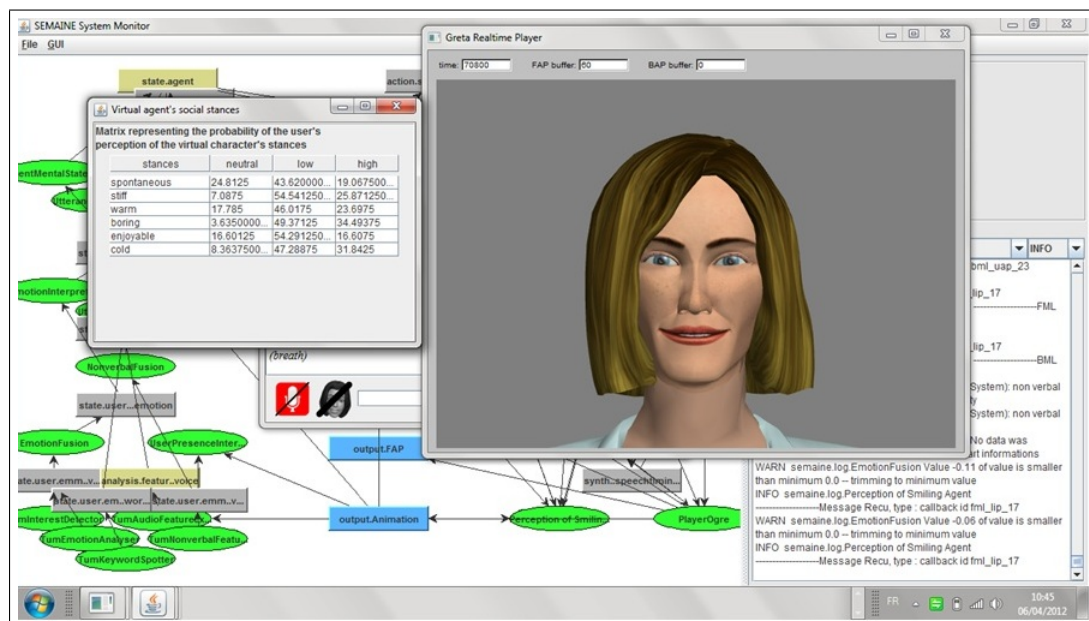
Fig. 10. Screenshot of the SEMAINE platform including the matrix of probabilities that represent the potential user's perception of the ECA Poppy

tion step, the potentially perceived interpersonal stances are computed by averaging between the current values of the matrix and the computed values resulting from the situation. The following example illustrates the consequences of this approach. If several successive sentences that reflect a positive emotion—which would normally lead a user to suspect a positive interpersonal stance on the part of the ECA—are not accompanied by the display of an amused smile then this would lead, in terms of the model, to successive decreasing of a user's perception of the ECA's positive interpersonal stance. Through this hypothesis, we suppose that the expression or non-expression of a smile has the same impact on the user's perception whenever it occurs during the interaction whether the smile is expressed or not (e.g. beginning, middle or end of the interaction).

In the next section, we propose an evaluation studies to try to evaluate the model during virtual agent-user interaction.

5.2. Validation of the model of the user's perception in virtual agent-user interaction

The proposed model of user perception of smiling ECA interpersonal stances requires validation within the context of a social interaction. It was constructed based on results from a specific scenario (saying a riddle) and with an assumption that the computed probabilities are cumulative during the interaction. In order to evaluate the model of user perceptions of virtual agent's interpersonal stances, we performed an experiment detailed in the remainder of this section.

Method. Participants were asked to interact with the two virtual agents (Poppy and Obadiah, Figure 7) using the SEMAINE platform (Section 3). The dialog module of SEMAINE was modified to have exactly the same dialog behavior (i.e. same repertoire of questions and responses to the users) for the two virtual agents. To measure the impact of the smiling behavior, we implemented two versions of the virtual agents:

— a *non-smiling version* in which the virtual agents do not express any smiles;
— *a smiling version* in which the virtual agents display both amused and polite smiles in accordance with the communicative intentions associated with these smiles (Section 3).

Participants interacted with the two agents twice (the two smile conditions) for 3 minutes, creating a 2 (smile condition) x 2 (appearance) within participants design. The user looked into a tele-prompter, which consists of a semi-silvered screen at 45 degree to the vertical, with a horizontal computer screen below it, and a battery of cameras behind it[13]. The user saw the face of the ECA (Figure 11). Participants were
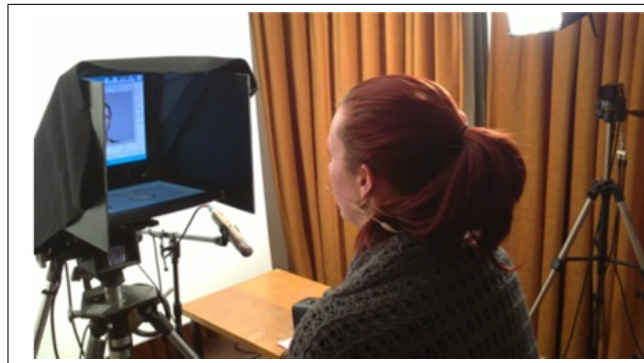

Fig. 11.   Setup of the experiment for the evaluation of the model of perception

recruited from the School of Psychology at Queen's University Belfast (N=15 of which 10 were female and 5 were male) and the majority were Northern Irish[14].

---

[13]Note that the tele-prompter is not necessary for human-virtual agents interactions. However, the study was conducted using a SEMAINE set-up that was designed to allow comparisons between human-human interactions (in which the tele-prompter is needed) and human-agents interactions [McKeown et al. 2012].
[14]Compared to this study, the culture of the participants in the previous studies described in the paper was mainly French. This culture difference should not be an issue since no difference among Western cultures has been reported for smiling behavior [Jiurgens et al. 2013].

Each participant engaged in four sequential three minute interactions: with the smiling virtual agent *Poppy*, the smiling virtual agent *Obadiah*, the non-smiling agent *Obadiah*, and the non-smiling agent *Poppy*. The order of smiling and non-smiling interactions, and of agents were counterbalanced to control for order effects. After each interaction, we asked participants to complete a questionnaire concerning their perception of the virtual agent's social stances, in particular: *warm, spontaneous, enjoyable, boring*, and *cold*. Additionally, the participants indicated their perception of the naturalness of the virtual agent. Participants rated these variables on a Likert scale of 10 points by answering a series of questions (*e.g.*, Did you find the character was warm?). In total, 60 user perceptions were collected using the questionnaire (30 of the virtual agent *Poppy* and 30 of the virtual agent *Obadiah*).

Results. To evaluate the accuracy of our model in estimating the user perceptions of virtual agent's social stances, we have compared the model output at the end of the interaction to participants' responses to the questionnaire. The output of the model corresponds to a matrix of probabilities (Figure 9). For each stance (*i.e.* each line of the matrix), we assume that the result of the computational model is the category with the highest probability. For instance, if the output of the model is the matrix illustrated in Figure 9, the result is that the ECA is perceived neutrally warm, low spontaneous and enjoyable, and high boring and cold.

The output of the model has been first evaluated based on a statistical analysis. For the warm stance, the model outputs for 45 of the conversations were low and the other 15 were neutral. We ran a one-way ANOVA to compare the user ratings for the low warm conversations compared to the neutral warm conversations. The analyses revealed that the low warm conversation (M=7.2) were rated statistically significantly higher (F=13.25, p=.001) than the neutral warm conversations (M=4.8). For the boring stance, the model outputs for 25 of the conversations were high and the other 35 were low. The analyses of a one-way ANOVA revealed a statistically significant difference (F=4.43, p=.04) showing that the high boring conversations (M=5.36) were also rated higher by the participants than the low boring conversations (M=3.94). Likewise, a one-way ANOVA comparing the high cold conversations (N=33, M=4.39) and the low cold conversations (N=27, M=2.11) showed a statistically significant difference, F=12.05, p=.001. The model outputs for all 60 conversations for fun and spontaneous were low and consequently no comparison of categories can be computed. These statistical results show that the conversations that the model stated highest in warmth, cold and boring also had a higher mean rating from the participants.

However, these results do not validate the categories (neutral, low or high) estimated by the model. For this purpose, we have to define a mapping between the users' ratings and the *neutral*, *low*, and *high* categories of the model. Indeed, the questionnaire produced users ratings in the form of values between 1 and 10, whereas, the output of the model produced the discrete categories *neutral*, *low*, and *high*. To compare the responses to the questionnaire and the outputs of our model, we have assumed that a user rating of 1 corresponds to the *neutral* category since this answer was labeled with "Not at all" on the questionnaire submitted to the participants. The mapping between the users' ratings in [2, 10] and the categories *low* and *high* is more questionable given the "fuzziness" of these categories. In order to consider this fuzziness, we evaluated the accuracy of the model by considering the 4 different mappings illustrated Table 4.

For each mapping, to report the accuracy of the computational model to estimate the user perception of the ECA's stances, we computed the true positive rate (*i.e.* number of good predictions of the computational model) and the false positive rate (*i.e.* number of bad predictions of the computational model). The results are plotted in the ROC (Receiver Operating Characteristic) space (Figure 12). The mapping with id.

Table IV. The different mappings between the users' ratings and the categories *low* and *high* considered in the evaluation of the computational model

Mapping id.	*low* category	*high* category
1.	users' rating in $[2, 7]$	users' rating in $[8, 10]$
2.	users' rating in $[2, 6]$	users' rating in $[7, 10]$
3.	users' rating in $[2, 5]$	users' rating in $[6, 10]$
4.	users' rating in $[2, 4]$	users' rating in $[5, 10]$
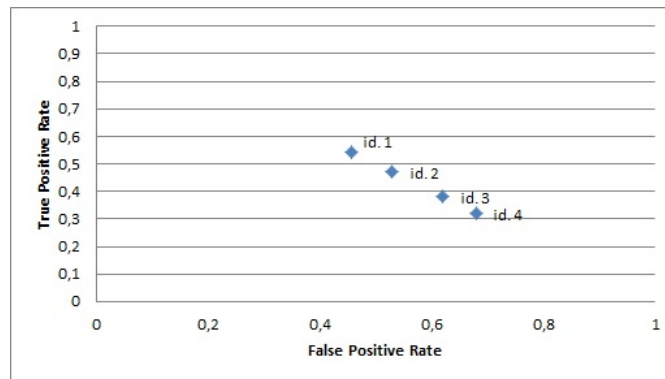
Fig. 12.   ROC space and plot of the four predictions of the computational model for each considered mapping (Table 4).
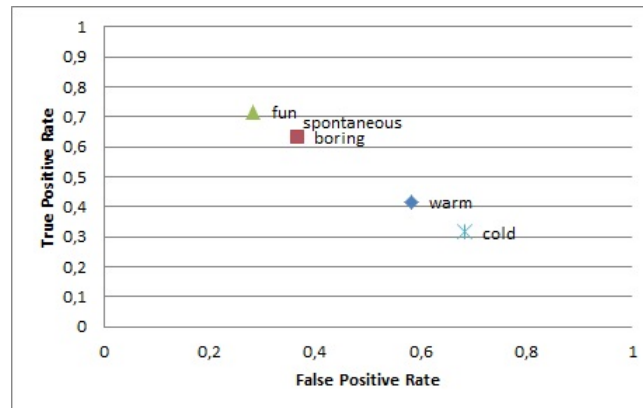

Fig. 13.   ROC space and plot of the five predictions of the computational model for each stance for the mapping id. 1 (Table 4). Spontaneous and boring appear in the same place in the figure.

1 had the best predictive power among all the explored mappings (Figure 12). This mapping had user ratings in the $[2, 7]$ interval corresponding to the *low* category and the user ratings in the [8,10] interval corresponding to the *high* category; at these levels the computational model estimated the user perception of an ECA's stance with a probability equal to 54%. Considering only this mapping (id. 1) the produced ROC space for each interpersonal stance is illustrated in Figure 13. The model's capacity to estimate user perceptions of an ECA depends on the stance. The probability of accurately estimating user perceptions of the fun, spontaneous, and boring stances of the ECA is superior to 50%. However, the computational model did not provide good prediction for the warm and cold stances. A comparison of the accuracy

of the model depending on the appearance of the ECA did not reveal strong differences.

Finally, concerning the naturalness of the ECA, a repeated-measures ANOVA revealed a statistical trend ($F = 3.37$, $p = .088$) showing that the smiling conditions are rated higher in naturalness than the non-smiling conditions.

Discussion. The proposed model to predict the user perceptions of ECA's social stances has been partly validated. The statistical analysis has shown that the model rated some conversations in a similar way to users. However, the lack of variability in the outputs of the model for the fun and spontaneous stances did not enable us to statistically evaluate the model for these stances. To complete this analysis, the accuracy of the model outputs was evaluated, that is, the capacity of the model to produce the appropriate categories (*neutral*, *low* and *high*). One issue is the comparison of user ratings (values ranging from 1 to 10) to the fuzzy categories of the model (*neutral*, *low* and *high*). To overcome this problem, the accuracy of the model was evaluated with a variety of different mappings. The mapping leading to the best accuracy corresponds to an extensive *low* category (in this mapping, users ratings from 2 to 7 were considered in the low category). Based on this mapping, the results showed that the model may be used to estimate the users perceptions of certain stances (fun, spontaneous, and boring). The selected mapping is particularly aligned to these stances since most of the outputs of the computational model corresponds to the *low* category. However, a model that provides always the same output whatever the ECA's smiling behavior, is not particularly useful. Our model requires improvement for the *fun* and *spontaneous* stances. Fine-grained categories could be considered to catch more subtle variations in user perceptions of these ECA stances. Concerning the warm and cold stances, the model did not accurately predict the user perceptions of these stances. Given the results of the statistical analysis of these stances, the proposed computational model could be improved by considering other categories. Indeed, the neutral, low, and high categories did not seem to reflect the variation in users' ratings for these stances. Concerning the appearance of the ECA, the model accuracy in predicting stances for the virtual agent *Poppy* or *Obadiah* is relatively similar.

Given that the accuracy of the model was computed based on a particular mapping between users' ratings and categories suitable to the obtained data, an evaluation should be carried to ensure that the selected mapping is similarly aligned in other human-agent conversations. To overcome this categorization problem, in a future evaluation of the model, the outputs of the computational model will be modified to directly consider the numerical values between 1 and 5 and the associated probabilities. In this case, the model accuracy could be evaluated by a computation of the correlations between the users ratings and the numerical outputs of the model. Then, after a validation of the model, the outputs could be transformed to categories. To refine the model, we aim to understand the variation within the participant ratings. In this study we have focused on smiles, however, others elements may influence the perception of ECA's stances, such as the appearance of the ECA (its gender, its age, its ethnicity, etc.) or the course of the conversation. For this purpose, an initial questionnaire to collect the first impression of the ECA could be defined. The questionnaire after each interaction could be extended to questions concerning the flow of the conversation or the topics discussed.

The model has been integrated and tested in a particular human-agent context—a user involved in a small-talk conversation with a virtual agent. In this context, the virtual agent did not embody a specific social role. In a task-oriented application, such as in a virtual learning class, the context (*e.g.* the role of the virtual agent and of the user) may influence the perception of the ECA's stances. Moreover, the ECA may have

to express specific stances required by its social role and the course of the interaction. For instance, a virtual tutor should not express a fun stance when user performance on an educational task decreases. Deviation from expected social stances of this nature may lead to particular user feelings (e.g. confusion) and may impact his performances in task achievement. The proposed computational model should therefore been evaluated in a task-oriented domain. In task-oriented contexts, unexpected ECA smiling behavior may have a stronger impact on user perception. The model could be enriched by considering both the effects of smiling behaviors on the perceived stances but also the possible effects of its smiling behavior on other elements such as the engagement or performances of the user.

6. CONCLUSION

In conclusion, the work presented in this paper aimed at endowing an ECA with a "smile focused theory of mind" to automatically infer, during the interaction, the potential user's perception of its social stances depending on its smiling behavior. In order to give the capability to a virtual agent to display different smiling behaviors, we constructed a lexicon of virtual agent smiles including different smile types (amused, embarrassed and polite). The methodology used to define the morphological and dynamic characteristics of the smiles is based on the design, by the users themselves, of the different types of smile. The proposed algorithm used to extract the smiles characteristics from the collected data has the advantage to capture one-to-many correspondence between smile types (amused, embarrassed and polite) and facial expressions, offering a variability in ECA's smile expressions. The resulting smiles, evaluated in context through a perceptive study, are fairly consistent with the studies on human smiles, suggesting that users apply common knowledge on human smiles to virtual agents. The different smiles have been integrated in an ECA. Note that in this article, we have focused on three types of smile largely studied in interpersonal interaction. In a more exploratory work, we have used unsupervised methods on the user created smiles to explore new types of smile without considering *a priori* typology of smile types [Ochs et al. 2015].

These smiles may accompany an ECA goal to express a certain communicative intention or stance. Given the impact of its smile expressions on its perceived stances, an ECA may decide to express or not express a smile depending on the stance it wants to convey. In order to give the ECA the capability to identify the potential the impact of its smiling behavior, we developed a probabilistic model that estimates the user perception of the virtual agent's social stances depending on its expressed smiles. The model has been constructed based on a corpus of user perceptions of smiling and non-smiling virtual agents with different appearances. An experiment using real human-virtual agent interaction provided some validation of the proposed model. The evaluation highlighted the difficulty in validating the fuzzy categories (neutral, low and high) used to characterize the intensity of the perceived stances. Moreover, some stances were not well predicted by the computational model. Globally, the performance of the model remains only average. The performance of the model may be explained by the fact that the perceived stances of the ECA may be influenced not only by the smiling behavior but also by others elements such as the other non-verbal expressions (e.g. head nods) and the course of the conversation (e.g., by out of place responses of the agent that may happen during the course of the interaction). To develop a more complete theory of mind model of the ECA's stances, the different elements that may impact the perception of stances would need to be taken into account—this is a task that will remain difficult with current technologies for some time to come. Nevertheless, the proposed model represents a first step in the construction of such a model.

One main issue of the present work is its generic aspect. In fact, the presented studies have been performed using particular ECAs with specific appearances. We cannot conclude that the resulting model may be applied to other virtual agents. Indeed, research has highlighted the strong impact that a virtual agent's appearance may have on a user during an interaction [Baylor 2009]. The different parameters of a virtual agent's appearance that may influence user perception of smiles and social stances should be identified. For this purpose, the proposed studies could be replicated including a variety of virtual faces to assess results with different virtual agent appearances. The proposed model of smiling virtual agent is certainly dependent on the virtual agent's appearance but the proposed methodology to construct a model of smiling virtual agent remains independent from the virtual agent.

The smiling virtual ECAs have been evaluated in the context of a small-talk conversation with humans. In this context, only the capacity of the ECAs to convey different stances through smiles has been evaluated. In a task-oriented context, some ECA's stances could be more appropriate than others. The computational model should be evaluated in different task-oriented contexts by considering the influence of stances expressions on the performances of the users in task achievement. In this way, the most appropriate stances to express given the context of the interaction for optimal performances of the user could be identified.

The presented studies have been conducted in the context of Western Culture. The perception and the production of smiles are highly culturally dependent. As shown in [Thibault et al. 2012], even the nuances of the different types of smile expression (e.g. the AU6 to distinguish amused and polite smile) are not universal. In [Beaupré and Hess 2003], the authors highlight the different perceptions of smiles depending on the cultural group of the perceived smiling face (European, Asian or African). The social rules on the conditions in which a smile is expected, and consequently the stances expressed through smiling behavior, also depend strongly on the culture [Hess et al. 2002]. These cultural differences in smiling behavior highlight the importance to consider the resulting smiling model of the studies in the light of the considered culture of the participants influencing both the creation of smiles and the associated perceived stances.

ACKNOWLEDGMENTS

REFERENCES

Zara Ambadar, , Jeffrey Cohn, and Lawrence Ian Reed. 2009. All Smiles are Not Created Equal: Morphology and Timing of Smiles Perceived as Amused, Polite, and Embarrassed/Nervous. *Journal of Nonverbal Behavior* 17-34 (2009), 238–252.

Amy L. Baylor. 2009. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3559–3565.

Martin G Beaupré and Ursula Hess. 2003. In my mind, we all smile: A case of in-group favoritism. *Journal of Experimental Social Psychology* 39, 4 (2003), 371–377.

Michael J Bernstein, Donald F Sacco, Christina M Brown, Steven G Young, and Heather M Claypool. 2010. A preference for genuine smiles following social exclusion. *Journal of Experimental Social Psychology* 46, 1 (Jan. 2010), 196–199.

Elisabetta Bevacqua, Sylwia Hyniewska, and Catherine Pelachaud. 2010a. Positive influence of smile backchannels in ECAs. In *International Workshop on Interacting with ECAs as Virtual Characters, International Conference of Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Elisabetta Bevacqua, Ken Prepin, Radoslaw Niewiadomski, Etienne deSevin, and Catherine Pelachaud. 2010b. Artificial Companions in Society: Perspectives on the Present and Future. Chapter Greta: Towards an Interactive Conversational Virtual Companion, 143–156.

Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. 1984. *Classification and Regression Trees*. Chapman and Hall.

Judee K. Burgoon, Joseph A. Bonito, Paul Benjamin Lowry, Sean L. Humpherys, Gregory D. Moody, James E. Gaskin, and Justin Scott Giboney. 2016. Application of Expectancy Violations Theory to communication with and judgments about embodied agents during a decision-making task. *International Journal of Human-Computer Studies* 91 (2016), 24 – 36.

Justine Cassell. 2000. More Than Just Another Pretty Face: Embodied Conversational Interface Agents. *Commun. ACM* 43 (2000), 70–78.

Justine Cassell, Timothy Bickmore, Lee Campbell, Hannes Vilhjálmsson, and Hao Yan. 2001. More than just a pretty face: Conversational protocols and the affordances of embodiment. *Knowledge-based Systems* 14, 1-2 (2001), 55–64.

Doris M. Dehn and Susanne van Mulken. 2000. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies* 52, 1 (2000), 1–22.

Francine M. Deutsch, Dorothy LeBaron, and Maury March Fryer. 1987. What is in the Smile? *Psychology of Women Quarterly* 11 (1987).

Richard Dietz and Sebastiano Moruzzi. 2010. *Cuts and Clouds: Vagueness, its Nature, and its Logic*. Oxford University Press.

Guillaume Duchenne. 1990. *The Mechanism of Human Facial Expression*. Cambridge University Press.

Robin Ian MacDonald Dunbar. 2001. Theory of mind and the evolution of language. In *Approaches to the evolution of language*, J R Hurford, M Studdert-Kennedy, and C Knight (Eds.). Cambridge University Press, Cambridge, 92–110.

Joyce Edinger and Miles Patterson. 1983. Nonverbal involvement and social control. *Psychological Bulletin* 93 (1983), 30–56.

Paul Ekman. 1992. Facial expressions of emotion: New findings, new questions. *Psychological Science* 3 (1992), 34–38.

Paul Ekman. 2003. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences* 1000 (2003), 205–221.

Paul Ekman. 2009. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (ww norton & co ed.).

Paul Ekman and Wallace Friesen. 1975. *Unmasking the Face. A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Paul Ekman and Wallace Friesen. 1982. Felt, False, And Miserable Smiles. *Journal of Nonverbal Behavior* 6 (1982), 238–252.

Paul Ekman, Wallace Friesen, and Joseph Hager. 2002. *The facial action coding system*. Weidenfeld and Nicolson.

M. G. Frank, P. Ekman, and W. V. Friesen. 1993. Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology* 64 (1993), 83–93.

Phillip J Glenn. 2003. *Laughter in Interaction*. Cambridge University Press, Cambridge.

Karl Grammer and Elisabeth Oberzaucher. 2006. The Reconstruction of Facial Expressions in Embodied Systems. *ZiF : Mitteilungen* 2 (2006).

Jinny Harrigan and Dennis O'Connell. 1996. How do you look when feeling anxious? Facial displays of anxiety. *Personality and Individual Differences* 21 (1996), 205–212.

Ursula Hess, Martin G Beaupré, Nicole Cheung, and others. 2002. Who to whom and why–cultural differences and similarities in the function of smiles. *An empirical reflection on the smile* 4 (2002), 187.

Ursulla Hess, Sylvie Blairy, and Robert Kleck. 2000. The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior* 24 (2000), 275–283.

Uruslla Hess and Robert Kleck. 1990. Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology* 20, 5 (1990), 369–385.

Dirk Heylen. 2006. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics* 3, 3 (2006), 241–267.

Mohammed Hoque, Daniel McDuff, and Rosalind Picard. 2012. Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. *IEEE Transactions on Affective Computing* 3, 3 (2012), 323–334.

Mohammed Hoque, Louis-Philippe Morency, and Rosalind Picard. 2011. Are you friendly or just polite? - analysis of smiles in spontaneous face-to-face interactions. In *Proceedings of the international conference on Affective Computing and Intelligent Interaction (ACII)*. Springer-Verlag, Berlin, Heidelberg, 135–144.

Rebecca Jiurgens, Matthis Drolet, Ralph Pirow, Elisabeth Scheiner, and Julia Fischer. 2013. Encoding conditions affect recognition of vocally expressed emotions across cultures. *Frontiers in Psychology* 4, 111 (2013).

Dacher Keltner. 1995. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology* 68(3) (1995), 441–454.

Timothy Ketelaar, Bryan L Koenig, Daniel Gambacorta, Igor Dolgov, Daniel Hor, Jennifer Zarzosa, Cuauhtémoc Luna-Nevarez, Micki Klungle, and Lee Wells. 2012. Smiles as signals of lower status in football players and fashion models: Evidence that smiles are associated with lower dominance and lower prestige. *Evolutionary Psychology* 10, 3 (July 2012), 371–397.

Scott Fabius Kielsing. 2009. Stance: Sociolinguistic Perspectives. Oxford University Press, Oxford, Chapter Style as stance: Stance as the explanation for patterns of sociolinguistic variation, 171–194.

Mark L Knapp, Judith A Hall, and Terrence G Horgan. 2013. *Nonverbal Communication in Human Interaction*. Cengage Learning.

Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson, and Hannes Vilhjálmsson. 2006. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *Proceedings of the international conference on Intelligent Virtual Agents (IVA)*. Springer-Verlag Berlin, Heidelberg, 21–23.

Nicole Krämer, Stefan Kopp, Christian Becker-Asano, and Nicole Sommer. 2013. Smile and the world will smile with you-The effects of a virtual agent's smile on users'evaluation and behavior. *International Journal of Human-Computer Studies* 71, 3 (2013), 335–349.

Nicole C. Krämer. 2008. Social Effects of Virtual Assistants. A Review of Empirical Results with Regard to Communication. In *Proceedings of the international conference on Intelligent Virtual Agents (IVA)*. Springer-Verlag, Berlin, Heidelberg, 507–508.

Eva Krumhuber, Antony SR Manstead, and Arvid Kappas. 2007. Temporal aspects of facial displays in person and expression perception. The effects of smile dynamics, head tilt and gender. *Journal of Nonverbal Behavior* 31 (2007), 39–56.

Eva G Krumhuber and Antony SR Manstead. 2009. Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 6 (2009), 807–820.

Marianne LaFrance and Marvin A Hecht. 1995. Why smiles generate leniency. *Personality and Social Psychology Bulletin* 21, 3 (1995), 207–214.

Sing Lau. 1982. The effect of smiling on person perception. *Journal of Social Psychology* 117 (1982), 63–67.

Winter Mason and Siddharth Suri. 2011. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (June 2011), 1–21.

Richard E Mayer and C Scott DaPra. 2012. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied* 18, 3 (2012), 239.

Gary McKeown, Ian Sneddon, and W Curran. 2015. Gender Differences in the Perceptions of Genuine and Simulated Laughter and Amused Facial Expressions. *Emotion Review* 7, 1 (Jan. 2015), 30–38.

Gary McKeown, Michel Valstar, Rodie Cowie, Maya Pantic, and Marc Schroder. 2012. The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing* 3, 1 (2012), 5–17.

Gary J McKeown. 2013. The Analogical Peacock Hypothesis: The sexual selection of mind-reading and relational cognition in human communication. *Review of General Psychology* (2013).

Marc Mehu, Anthony Little, and Robin I.M. Dunbar. 2007. Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology* 5, 1-4 (2007), 133–146.

Monica Moore. 1985. Nonverbal courtship patterns in women. *Ethology and Sociobiology* 6 (1985), 237–247.

Radoslaw Niewiadomski, Sylwia Hyniewska, and Catherine Pelachaud. 2011. Constraint-Based Model for Synthesis of Multimodal Sequential Expressions of Emotions. *IEEE Transactions on Affective Computing* 2, 3 (2011), 134–146.

Radoslaw Niewiadomski and Catherine Pelachaud. 2007. Model of Facial Expressions Management for an Embodied Conversational Agent. In *Proceedings of the international conference on Affective Computing and Intelligent Interaction (ACII)*. 12–23.

Kristine L. Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoper. Virtual Environ.* 12, 5 (Oct. 2003), 481–494.

Magalie Ochs, Edwin Diday, and Filipe Afonso. 2015. From the symbolic analysis of virtual faces to a smiles machine. (2015).

Magalie Ochs, Radoslaw Niewiadmoski, and Catherine Pelachaud. 2010. How a virtual agent should smile? Morphological and dynamic characteristics of virtual agent's smiles. In *Proceedings of the international conference on Intelligent Virtual Agents (IVA)*. Springer Berlin Heidelberg, 427–440.

Magalie Ochs, Radoslaw Niewiadomski, Paul Brunet, and Catherine Pelachaud. 2012. Smiling Virtual Agent in Social Context. *Cognitive Processing, Special Issue on"Social Agents"* 13, 22 (2012), 519–532.

Emma Otta, Fabiana Follador E Abrosio, and Rachel Leneberg Hoshino. 1996. Reading a smiling face: messages conveyed by various forms of smiling. *Perceptual and motor skills* 82, 3 Pt 2 (June 1996), 1111–1121.

John ODoherty, Joel Winston, Hugo Critchley, David Perrett, D Michael Burt, and Raymond J Dolan. 2003. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41, 2 (2003), 147–155.

David Pardo, Beatriz L Mencia, Álvaro H Trapote, and Luis Hernández. 2009. Non-verbal communication strategies to improve robustness in dialogue systems: a comparative study. *Journal on Multimodal User Interfaces* 3, 4 (2009), 285–297.

Iabella Poggi and Roberto Chirico. 1998. The meaning of smile. In *Oralité, gestualité, communication multimodale, interaction*. 159–164.

Isabella Poggi and Catherine Pelachaud. 2000. Emotional Meaning and Expression in Animated Faces. In *Affective Interactions*, Ana Paiva (Ed.). Lecture Notes in Computer Science, Vol. 1814. Springer Berlin / Heidelberg, 182–195.

Ricco Rakotomalala. 2005. TANAGRA : un logiciel gratuit pour l'enseignement et la recherche. In *Extraction et Gestion des Connaissances (EGC)*. 697–702.

Mathias Rehm and Elisabeth André. 2005. Catch me if you can ? Exploring lying agents in social settings. In *Proceedings of the international conference of Autonomous Agents and Multi-Agent Systems (AAMAS)*. Academic Press Inc, 937–944.

Harry T Reis, Ilona McDougal Wilson, Carla Monestere, Stuart Bernstein, Kelly Clark, Edward Seidl, Michelle Franco, Ezia Gioioso, Lori Freeman, and Kimberly Radoane. 1990. What is smiling is beautiful and good. *European Journal of Social Psychology* 20 (1990), 259–267.

Laurel D Riek and Peter Robinson. 2011. Challenges and Opportunities in Building Socially Intelligent Machines. *Signal Processing Magazine, IEEE* 28, 3 (May 2011), 146–149.

Klaus Scherer. 2005. What are emotions? And how can they be measured? *Social Science Information* 44, 4 (Dec. 2005), 695–729.

Marc Schröder. 2010. The SEMAINE API: Towards a Standards-Based Framework for Building Emotion-Oriented Systems. *Advances in Human-Computer Interaction* vol. 2010, Article ID 319406 (2010), 21 pages.

Marc Schröder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Björn Schuller, Etienne de Sevin, Michel F Valstar, and Martin Wöllmer. 2012. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3, 2 (2012), 165–183.

Ian Sneddon, Margaret McRorie, and Gary McKeown. 2012. The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing* 3, 1 (2012), 32–41.

Jacalyn Snodgrass. 1992. *Judgment of feeling states from facial behavior: A bottom-up approach*. Ph.D. Dissertation. University of British Columbia.

Chunyang Su, Shifei Ding, Weikuan Jia, Xin Wang, and Xinzheng Xu. 2008. Some Progress of Supervised Learning. In *Proceedings of the international conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications - with Aspects of Artificial Intelligence (ICIC '08)*. Springer-Verlag, Berlin, Heidelberg, 661–666.

Emmanuel Tanguy. 2006. *Emotions: the art of communication applied to virtual actors*. Ph.D. Dissertation. Department of Computer Science, University of Bath, England.

G Theonas, Dave Hobbs, and Dimitrios Rigas. 2008. Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments. *International Journal of Virtual Reality* 7 (2008), 31–44.

Pascal Thibault, Manon Levesque, Pierre Gosselin, and Ursula Hess. 2012. The Duchenne marker is not a universal signal of smile authenticity–but it can be learned! *Social Psychology* (2012).

Michael Tomasello. 1999. *The cultural origins of human cognition*. Harvard University Press, Cambridge, MA.

Ning Wang, W. Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. 2005. The Politeness Effect: Pedagogical Agents and Learning Gains. In *Proceedings of the 2005 conference*

on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology. IOS Press, Amsterdam, The Netherlands, 686–693.

Joseph Weizenbaum. 1966. ELIZA-a computer program for the study of natural language communication between man and machine. *Communication of the ACM* 9, 1 (1966), 36–45.

Andrew Whiten, , and Richard Byrne. 1997. *Machiavellian intelligence II*. Cambridge University Press, Cambridge, UK.

Man-Ching Yuen, I. King, and Kwong-Sak Leung. 2011. A Survey of Crowdsourcing Systems. In *Proceedings of the IEEE International conference on Social Computing (SocialCom)*. 766 –773.