"I Know That You Know How I Feel": Behavioral and Physiological Signals Demonstrate Emotional Attunement While Interacting with a Computer Simulating Emotional Intelligence

Stefania Balzarotti · Luca Piccini · Giuseppe Andreoni · Rita Ciceri

Published online: 13 April 2014

Introduction

It is a quite common experience to feel frustrated or even angry when interacting with our computers. When something goes wrong (e.g., important files are lost, the computer

S. Balzarotti · R. Ciceri (✉)
Laboratory of Communication Psychology, Catholic University of the Sacred Heart, Largo Gemelli, 1, 20123 Milan, Italy
e-mail: maria.ciceri@unicatt.it

L. Piccini · G. Andreoni
SensibiLab, Department of INDACO, University Politecnico di Milano, Lecco, Italy

becomes very slow, etc.), sometimes people swear, yell at their PC, hit the keyboard or the mouse, that is, they express their emotions as if computers were social agents that can be scolded, threatened or implored to make them work better (Surakka and Vanhala 2011).

In order to improve and facilitate human–computer interactions (HCI), *Affective Computing* researchers have turned their attention to the emotions experienced and expressed by human users, in the effort to implement *emotionally intelligent* interfaces able to automatically decode these emotions and respond properly (e.g., Gratch and Marsella 2007; Höök 1998; Lisetti et al. 2003; Picard 1997; Picard et al. 2001). Researchers have thus started questioning which principles and features that govern human–human (HH) communication should be applied to the design of HC affective dialogs (Cappella and Pelachaud 2001; Höök 1998). Also, researchers have started analyzing the emotional responses exhibited by users within HC interactions in order to identify the signals that are most relevant to (1) recognize the users' emotions (e.g., Bailenson et al. 2008) and (2) implement expressive feedback with virtual agents such as Embodied Conversational Agents (e.g., Caridakis et al. 2006).

The main purpose of the present paper is to offer a contribution to address these questions. In particular, the paper will advance the hypothesis that a useful principle to model believable affective HC interactions is the psychological construct of *emotional attunement*. Communicative partners are "attuned" when they acknowledge each other as intentional agents, and they adjust their behavior one to another (Ciceri and Biassoni 2006; Searle 1998; Siegman and Feldstein 1979). In the present study, we examine the emotional responses of users when they are told that the computer is an intentional agent able to understand their emotions: Do users exhibit a larger amount of emotional behaviors in this condition than when interacting with a standard PC? And is the HC interaction perceived as more effective? Second, we examine whether users adjust their emotional responses to the computer's "behavior", that is, to the type of activity run by the computer.

The Attunement Process: Awareness and Adjustment

The attunement process has been extensively studied in HH communicative interactions (e.g., Giles et al. 1992, 2001; Siegman and Feldstein 1979). Studies and theories have highlighted two main components.

First, HH communicative interactions imply the mutual acknowledgement of partners who perceive each other as agents characterized by intentionality and goal-oriented agency. Said in other words, individuals usually attribute internal states—such as intentions, thoughts and emotions—to their interlocutors. Cooperation (Grice 1989) and reciprocal ratification (Goffman 1967) are preconditions of any communicative interaction between two or more partners. In a similar way, the existence of a communicative agreement between the actors is assumed as a crucial factor for the achievement of a common goal by Searle's WE-Intention theory (1998).

As a second requirement, human communication implies the mutual adjustment and coordination of the actors' behavior. Communication theorists have defined "attunement" as a set of nonverbal behavioral units through which the speakers are able to manage, maintain and coordinate their communicative interaction being "on the same wavelength" (Siegman and Feldstein 1979). For instance, according to the communication accommodation theory (CAT), partners tend to *accommodate* one to another by adjusting their behavior to the context, topic of discourse, partner characteristics, purpose of the interaction and roles assigned to participants in that given context (Giles et al. 2001). Within this approach, attunement consists of a large number of both linguistic and nonverbal signals—such as length of utterances and pauses, voice volume, pitch, speech rate, etc.—through which one's own communicative style is tailored to the conversational style of the partner, in order to foster communication intelligibility and efficacy (Giles et al. 2001; Siegman and Feldstein 1979; Zerubavel 1981). It is worth noting that attunement and communicative synchrony characterize human individuals from birth: Starting from the microanalysis of face to face interactions in mother-infant dyads, Stern (1985) defined *affective attunement* as a primarily nonverbal mode of communication, consisting in the 'intermodal matching' of intensity, timing and shape of behavior on the basis of micro-dynamic modifications over time (Beebe et al. 2005; Trevarthen 1993).

Emotional Attunement: Emotion Nonverbal Behavior

Among the nonverbal behavioral units modulated to accomplish attunement—a key role is played by emotional signals. *Emotional attunement*—that can be defined as the attunement accomplished through the expression of emotional nonverbal signals—enables the synchronized flow of the interaction (Clark 1985; Clark and Wilkes-Gibbs 1986; Garrod and Anderson 1987; Giles et al. 1992). Emotion signaling processes (for instance, by means of facial displays) serve in fact important social functions and convey information about the individual's emotions, intentions and goals (e.g., Frijda and Mesquita 1994; Keltner and Haidt 1999; Van Kleef et al. 2010).

A bulk of studies have shown that individuals employ nonverbal expressions of emotions to communicate and infer cooperation in negotiation tasks within the context of either HH (Boone and Buck 2003; Krumhuber et al. 2007) and HC interactions (de Melo et al. 2010), suggesting that cooperators may be more emotionally expressive than non-cooperators (Schug et al. 2010). Another group of studies has shown that when individuals watch an emotional videotape with others (for instance, friends) who are either physically (i.e., in the same room) or implicitly (i.e., in another room) present, they exhibit more intense nonverbal expressions compared to situations when they watch the video alone (Chovil 1991; Fridlund 1991) or when the other partner is engaged in a different activity (e.g., answering a questionnaire; Hess et al. 1995). These findings suggest that high sociality contexts (i.e., individuals are aware of the presence of others) elicit strong motive to communicate (Chovil 1991).

Emotional Attunement in HC Dialogs

The issues considered above open interesting questions when applied to the context of HM interaction. Do users acknowledge computers as communicative partners? And do users emotionally attune to them? According to the well-known Media Equation, the answer should be "yes": People ascribe human characteristics—as for instance personality traits—to artificial systems, treating computers more like social partners rather than mere tools (Nass and Lee 2000; Reeves and Nass 1996). Thus, individuals are foremost social in their interactions, even when interacting with inanimate computers (Surakka and Vanhala 2011). A number of studies have brought evidence that users change their behavior and they adapt their dialog style—for instance speaking more carefully and less naturally—when aware of interacting with a machine (Oviatt and Adams 2000; Oviatt et al. 2005). Referring to the CAT, Darves and Oviatt (2002) have shown that humans adjust their behavior to converge with a computer partner.

What about attuning to an emotionally intelligent computer? A number of studies demonstrated that users talk more to anthropomorphic interfaces, are more likely to attribute intelligence to them, and rate as more pleasant interfaces presented with a realistic face (e.g., Brennan and Ohaeri 1994; King and Ohya 1995; Koda and Maes 1996). More recently, however, it has been suggested that the use of anthropomorphic (i.e., artificial agent with bodies and facial or vocal emotional expression very similar to the human ones) avatars may not be sufficient to involve the user in true affective interactions (Höök 2004). Klein et al. (2002) showed that users continued to interact with a computer that had caused their frustration significantly longer if they interacted with an affect-support agent designed to provide feedback on the user emotional state compared to a condition when the users' emotions were ignored. In another study (Axelrode and Hone 2005), participants played a word game under four different conditions: (1) they interacted with either a simulated system that appeared to act affectively (sending feedbacks to their emotional responses) or with a standard PC; (2) they were either told that the computer was a prototype system that may be able to understand their emotional expressions or they were not told anything about this ability. Participants displayed a higher number of positive nonverbal expressions when they were told that the system might respond to their emotions and when the system did in fact show affective feedbacks.

Several authors have thus advanced dyadic principles to model HC interactions: a "user-centered" approach (Höök 2004); intimacy (Cassel and Bickmore 2003); empathic engagement (Hall et al. 2005); contingent responsiveness (Cappella and Pelachaud 2001); and intersubjectivity (Cassell and Tartaro 2007). Although these principles are similar to the concept of emotional attunement, they capture alternatively only one of the two components of awareness and adjustment.

The Present Study

The present study represents a first attempt to structure an experimental protocol to investigate the emotional attunement between a human user and a (simulated) emotional intelligent computer, analyzing the previously mentioned components of the emotional attunement process within the context of HCI. In our study, participants interacted with a computer provided with an avatar that guided them across four different game-like activities.

We tested two main hypotheses. Our first hypothesis concerned mutual awareness, that is, if the user is aware that the artificial agent is able to understand his or her emotional responses and to change its action accordingly, he or she may participate in the interaction by exhibiting a higher number of emotional signals. In order to test this hypothesis, participants were assigned to two different conditions. In the *awareness condition*, they were asked to play with a simulated emotional-intelligent computer (Lisetti 2002; Kort et al. 2001): The avatar introduced itself as an agent able to monitor and understand human emotional reactions and—after each activity—commented upon the participants' emotions. By contrast, in the *control* condition, the avatar merely provided instructions about the video activities, without mentioning any ability to understand emotional responses. We expected that participants in the awareness condition would exhibit significantly more emotional behavioral signals than participants in the control condition, and that they would be more involved and physiologically activated.

Our second hypothesis concerned the adjustment of the emotional signals to the computer "behavior". In this study, the computer behavior consisted of four video emotional activities which differed according to the interactivity level (low vs. high). We expected that—regardless of the type of condition—the exhibition of emotional signals would differ according to the computer stimuli, increasing when the computer confronted participants with highly interactive activities. The computer game-like activities were created through the manipulation of game events in a similar way to previous studies in the emotional literature (e.g., Kappas and Pecchinenda 2000; Van Reekum et al. 2004). These studies have shown that computer games can be a reliable instrument to elicit users' emotional responses.

Two main factors concerning emotion research and theory lie at the basis of the methodological approach followed in this study. First, we considered emotional responses belonging to physiology, expressive behavior and subjective experience, since they represent the three main emotion response systems indicated by emotion theories and research (e.g., Buck 1985; Ekman 1992; Izard 1977; Levenson 2003; for a review see Mauss and Robinson 2009). Concerning physiological signals, it is worth noting that—within the paradigm of Affective Computing—the use of biosignals has been largely employed as useful to analyze the impact of artificial interfaces on the user's emotional state (e.g., Picard et al. 2001; Prendinger et al. 2003). Far less common—to our knowledge—it is the use of physiological responses to investigate the attunement process (e.g., van Bakel and Riksen-Walraven 2008), perhaps because these signals are not under the individual's intentional control. However, given their large use in the study of HM interactions, we decided to include physiological signals within our analysis, expecting that they might nonetheless have a role in discriminating attunement phenomena. In particular, we hypothesized that the participants' physiological arousal would increase proportionally to their engagement (Gilleade et al. 2005), which was expected to be higher when playing with a computer recognized as "emotional intelligent" and in highly interactive activities.

Second, concerning emotion nonverbal behavior, we adopted a micro-analysis of behavioral units in multiple expressive systems (i.e., face, gaze, posture, and voice; Ciceri and Balzarotti 2008). Considering the multiplicity of emotional signals that characterize human spontaneous interactions, we expected that a wide range of nonverbal signals used to express blended emotional responses—e.g., frustration, boredom, amusement, etc.—would be extremely relevant, whereas stereotyped patterns and prototypical full-face expression of few basic emotions would appear less often within our dataset (Kaiser and Wehrle 2001). As this study is meant as an initial attempt to experimentally investigate the attunement process between human and computer, the analysis of the emotional reactions within different response systems may represent a first step towards the extraction of different types of parameters in order to identify the most reliable and significant to HC interaction.

Method

Participants

The sample consisted of 24 undergraduates (age ranging from 20 to 26, $M = 23.1$; $SD = 1.87$, 50 % female) attending either humanistic ($N = 12$) or scientific faculties ($N = 12$). All participants were volunteers and did not receive any credit for their participation in the study.

Computer Activities

Four different activities (Table 1) were designed in order to simulate low versus high levels of interactivity as well as to elicit emotional reactions.

Table 1 Experimental computer activities

Activity	Manipulated events	Expected emotions	Interactive level
Web exploration	Preselected virtual pages	None—Neutral	Low
Boring game	Carrots appear always in the same position	Boredom Frustration	Low (repetitive task)
Enemy game	Five levels of difficulty Positive versus Negative bonus Losing a life Poisoned carrot	Amusement Frustration Surprise	High
Quiz game	Easy question Difficult question	Amusement Frustration Surprise	High (question/answer structure)

Website Exploration

The first activity consisted in the exploration a few pages of a university website regarding undergraduate programs and courses. This activity was expected to be emotionally neutral and to involve a low interaction level as the participant is merely required to explore a controlled number of virtual pages.

Boring Game

The participant moves a rabbit avatar on the screen and has to collect a high number of carrots (50). Carrots appear one by one and always in the same positions, hence creating a repetitive and low-interactive task. This activity was thus expected to elicit negative emotions, such as boredom and frustration.

Enemy Game

In the same way as in the previous game, the participant controls a rabbit character that has to collect carrots. This activity, however, is highly interactive because the rabbit also needs to escape from an enemy. There are five levels of increasing difficulty. The player wins points for every carrot collected and every level successfully completed. Different events in the game are expected to elicit both positive and negative emotional responses (i.e., amusement, surprise, frustration, etc.): positive versus negative bonuses, passing successfully to the next game level versus losing a life (being captured by the enemy or eating poisoned carrots). Positive (money won) or negative bonuses (e.g., more enemies) appear randomly.

Quiz Game

15 questions of general culture are presented to the participant who has to choose the right answer among four alternatives. The participant wins money for every correct answer and loses money when answering wrong. Questions were divided into two series: A very easy one is followed by a very difficult one that makes the almost all the participant losing the prize won. In the same way as the enemy game, the quiz game was expected to elicit both positive and negative emotions such as amusement versus frustration, as well as to be

highly interactive. Nevertheless, the quiz game has a very quick question/answer structure, similar to a conversational exchange.

Experimental Set-Up

The experimental setting included two high-resolution web cameras. One camera was placed in front of the participant to record his or her facial movements, gaze direction and posture changes. A second camera was placed behind the participant in order to record his or her actions on the screen. Physiological signals (Electrocardiogram, ECG; Electrodermal Activity, EDA) were taken using the BIOPAC System (BIOPAC Systems Inc., Goleta, CA, http://www.biopac.com, 2004). Finally, a high quality microphone was used to record vocal reports. Physiological signals were synchronized through the use of an external trigger, which drove the BIOPAC system. The trigger was generated by the main computer, which also stored the video and audio data. The physiological signals were collected by a second PC connected to the BIOPAC.

Procedure

On arrival, participants were seated in a well-lit room. First, physiological sensors were attached explaining their functions. Electrodes were placed in standard positions (on wrists and ankle for the ECG, on second phalanges of the first and third finger of the non-dominant hand for the EDA). Participants were also informed that they would be videotaped and their prior consent was asked for the treatment of personal data for research purposes. Participants were then asked to start the computer session clicking on an icon placed on the desktop. They were told that they would have received all further instructions henceforth. They were then left alone in the room with the computer and, as the session started, an avatar (Baldi, CSLU Toolkit, http://cslu.cse.ogi.edu/toolkit/) accompanied by a computerized Italian speaking voice introduced them to the different kinds of computer activities through a 2 min brief discourse. At the end of each activity, the avatar spoke to the participants, commenting upon their performance and introducing the next game. The order of the activities was counterbalanced.

Participants received different instructions and information by the avatar determined by random assignment to one of the two different conditions. In the *awareness condition* (AW), participants were exposed to a simulated emotionally-intelligent computer (Kort et al. 2001; Lisetti 2002), which appeared to be automatic to the participants but it was actually controlled by an out-of-sight experimenter. The avatar claimed to be able to decode their emotional states and to adapt the tasks accordingly. For instance, the avatar used sentences like: “Thanks to the web cam and sensors I’m able to record your emotional reactions”, “You seem to be bored, so here it is a new game”, “You are in trouble: I’ll repeat the instructions to help you”. In the *control condition* (CT) the avatar guided the participants explaining the same activities but did not claim to be able to decode emotion nor used sentences where emotion was mentioned. Total duration of each session was of about 20 min (CT condition: $M = 1{,}391$ s, $SD = 201.68$; AW condition: $M = 1{,}270$ s, $SD = 136.68$).

Measures

Data were collected within three emotional response systems: subjective experience, nonverbal expressive behavior, and physiology.

Subjective Experience

At the end of the computer session, participants were asked to answer to a questionnaire. As a manipulation check (i.e., to test the efficacy of the simulated AW condition) participants were asked to rate on a seven point Likert scale the extent to which the computer had been able to (1) understand their emotions, (2) modify the task accordingly, and (3) provide instructions. On the same rating scale, they were also asked to judge how much the interaction with the computer had been clear and funny. Second, participants had to judge on a seven point Likert scale how much they had been involved by the four activities. Finally, they were asked to rate each computer activity according to three positive (amusing, pleasant, surprising; mean $\alpha = .77$) and three negative (boring, frustrating, annoying; mean $\alpha = .79$) emotional labels.

Nonverbal Behavior

All video tapes were coded *frame by frame* (25 fps) using The Observer 7.0 NOLDUS® by three coders who were unaware of the experimental conditions. This behavioral micro-analysis was based on the Behavioral Coding System (BCS; Ciceri and Balzarotti 2008), which included four macro-categories.

1. *Face*: The fundamental muscle movements that comprise Facial Action Coding System (Ekman and Frisen 1978) were selected. We considered action units relating to the upper face (AU 1, 2, 4, 5, 6, 7), lower face (AU 9, 10, 12, 14, 15, 16, 17, 25) and lip movements (AU 18, 20, 23, 24, 28 and AD 19, 30, 32, 33, 34, 37).
2. *Gaze direction:* We considered if the participant looked at the screen or away (e.g., at the keyboard, around the room, at himself/herself).
3. *Posture and head*: Concerning posture, behavioral units of moving near to/far from the screen were considered (approach vs. withdrawal), whereas concerning head movements, behavioral units were included such as head forward, backward, turned and tilted.
4. *Vocal behavior:* It was recorded each time the participant was speaking (verbal) or used other kind of vocalizations (grumbling, no-words etc.). Vocal behavior and face movements were mutual exclusive.

A total number of 40 categories were scored. Average inter-rater reliabilities were calculated (Cohen's kappa $= .89$). Frequency rates (number of occurrences/min) were extracted and computed for each macro-category (mean score). Moreover, to control for individual differences, change scores were computed by subtracting the rates extracted in the first 2 min of the session (baseline) to the rates obtained in each of the four activities. A total of 34,783 frames were analyzed for the CT condition and 31,753 for the AW condition.

Physiology

The ECG signal was filtered with a band-pass filter between 5 and 32 Hz in order to increase the signal to noise ratio of the QRS complex. A standard algorithm based on the wide-used Pam-Tompkins and already validated (http://www.physionet.org/) was used in order to extract heart rate (HR) and remove possible artifacts. The EDA signal was filtered with a 10 Hz low-pass filter an then down-sampled in order to increase the Signal-to-Noise ratio. Comparing EDA traces with videos allowed us to reject possible false responses caused by physical movement. Physiological responses were estimated through the same features used in previous literature concerning the investigation and identification of emotion through biosignals (Lisetti et al. 2003; Picard et al. 2001). The following six features were examined: (1) the means and (2) standard deviations of the selected signals, the absolute value of the first difference

$$\delta_{X,J} = \frac{1}{N_J - 1} \sum_{n-1}^{N_J - 1} |X_{n+1} - X_n| \tag{3}$$

the absolute value of the second difference

$$\lambda_{X,J} = \frac{1}{N_J - 1} \sum_{n-1}^{N_J - 2} |X_{n+2} - X_n|, \tag{4}$$

and the normalized first and second difference

$$\widetilde{\delta}_{X,J} = \frac{\delta_{X,J}}{\sigma_{X,J}}, \tag{5}$$

$$\widetilde{\gamma}_{X,J} = \frac{\gamma_{X,J}}{\sigma_{X,J}}. \tag{6}$$

All of these indexes were computed for EDA, whilst only the mean was extracted from the heart rate in the four different activities. HR and the features were extracted using custom software running on MATLAB ® (Mathworks Inc., Massachusetts, USA). In order to reduce the normal and natural variability that affects the physiological signals, all the synthetic values were normalized in respect of the initial basal stage, as explained in the analysis on nonverbal behavior previously described.

Results

Subjective Experience

Means and standard deviations are displayed in Table 2. *T* tests revealed that participants in the AW condition judged the avatar more capable to understand their emotions, $t(22) = 3.12$; $p < .01$, and to modify the task accordingly, $t(22) = 2.20$; $p < .05$) than participants in the control condition. Moreover, participants in the AW condition judged that the interaction with the computer was clearer, $t(22) = 2.24$; $p < .05$, and funnier, $t(22) = 2.80$; $p < .05$) than participants in the CT condition.

Second, answers about the perceived level of involvement were submitted to a mixed ANOVA (4 activities × 2 conditions). Results showed a significant main effect of type of activity, $F(3,66) = 26.23$, $p < .001$, $\eta^2 = .54$, and a significant interaction effect, $F(3,66) = 2.90$, $p < .05$, $\eta^2 = .12$. Univariate analysis revealed that AW participants were more involved by the quiz game $t(22) = 2.26$, $p < .05$, and less involved by the boring game, $t(22) = 2.44$, $p < .05$, than CT participants. Overall, the quiz and the enemy game were more involving than the boring game and the web exploration. In a similar way, positive and negative emotional labels were submitted to mixed ANOVAs (4 activities × 2 conditions). Results showed a significant main effect of the type of activity for both positive, $F(3,66) = 24.01$, $p < .001$, $\eta^2 = .52$, and negative emotions, $F(3,66) = 6.69$, $p < .01$, $\eta^2 = .23$, but no significant differences emerged between the two conditions. The

Table 2 Subjective experience: means and standard deviations of perceived computer abilities, level of involvement, positive and negative emotions

	Control *M* (*SD*)	Awareness *M* (*SD*)
Computer abilities		
Understand your emotions	2.33 (1.49)	4.08 (1.24)**
Modify the task accordingly	2.58 (1.62)	3.83 (1.12)*
Provide instructions	4.17 (1.58)	3.67 (1.87)
Level of involvement		
Web exploration	2.42 (1.79)	2.33 (1.89)
Boring game	4.17 (1.53)	2.83 (1.12)*
Enemy game	5.08 (1.38)	4.87 (1.23)
Quiz game	5.17 (1.27)	5.75 (1.29)*
Positive emotions		
Web exploration	6.92 (2.78)	6.50 (2.24)
Boring game	8.58 (2.93)	6.92 (2.11)
Enemy game	11.08 (3.52)	10.92 (2.90)
Quiz game	12.00 (2.70)	12.75 (3.11)
Negative emotions		
Web exploration	7.50 (2.51)	8.08 (1.93)
Boring game	9.17 (4.27)	11.33 (3.92)
Enemy game	8.50 (3.34)	9.75 (2.67)
Quiz game	6.08 (2.47)	7.08 (3.40)

* $p < .05$; ** $p < .01$

quiz and the enemy game were rated as more positive than the boring game and the web exploration; the boring game was rated as the most emotionally negative activity.

Nonverbal Behavior

We first tested that the type of condition had no effect on nonverbal behavior showed by participants during the initial 2 min (*baseline*) during which the avatar Baldi introduced and explained the activities. As expected, a MANOVA showed no difference in any of the behavioral macro-categories (*Wilks* $\lambda = 2.33$, $p > .05$), and thus we concluded that the random assignment of participants was successful.

Second, according to our hypotheses, we tested whether the two independent variables (i.e., awareness condition and type of activity) had a significant effect on the participants' nonverbal behavior: face, gaze, posture and vocal behavior (change scores). Means and standard deviations are shown in Table 3. The results of a mixed ANOVA (2 type of condition × 4 type of activity × 4 behavior) showed a main effect of type of activity, $F(3,66) = 80.05$, $p < .001$, $\eta^2 = .78$. Moreover, two significant second-level interactions emerged (Fig. 1): Type of Activity × Condition, $F(3,66) = 3.55$, $p < .05$, $\eta^2 = .14$, and Type of Activity × Macro-category effect, $F(9,198) = 41.81$, $p < .001$, $\eta^2 = .65$. The third-level interaction was also significant, $F(9,198) = 3.04$, $p < .05$, $\eta^2 = .12$. Univariate analysis revealed that participants in the AW condition showed a higher number of facial units in the quiz game, $t(22) = 2.72$, $p < .05$, and a lower number of behavioral units during the boring game than participants in the CT condition, $t(22) = 2.69$, $p < .05$ (Fig. 1a).

Table 3 Nonverbal behavior macro-categories: mean change scores (frequency rate) and standard deviations for type of activity and type of condition

	Face		Gaze		Posture		Voice	
	AW	CT	AW	CT	AW	CT	AW	CT
Web exploration	−1.62 (1.99)	−.89 (1.19)	−.21 (.33)	−.02 (.26)	−1.55 (1.42)	−1.74 (1.97)	−.57 (.55)	−.44 (.53)
Boring game	−2.89 (1.59)	−1.49 (.87)	−.12 (.33)	−.01 (.26)	−1.22 (1.12)	−1.24 (1.07)	−.44 (.78)	−.16 (.64)
Enemy game	−.84 (1.29)	−.86 (1.04)	.04 (.29)	.03 (.26)	−.94 (1.47)	−.65 (1.13)	1.60 (1.57)	.65 (1.10)
Quiz game	5.13 (1.78)	3.23 (1.63)	−.37 (.14)	−.36 (.22)	.95 (1.85)	.75 (1.73)	1.98 (1.94)	1.31 (1.57)

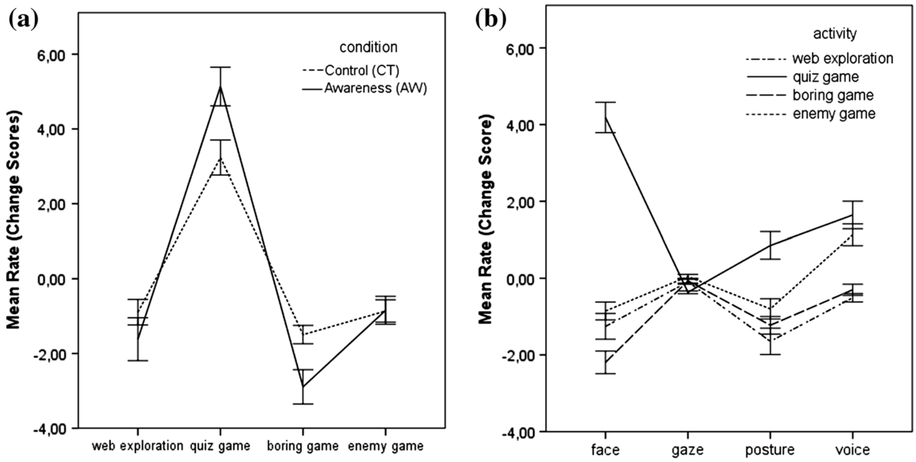

Fig. 1 Mean change scores in behavioral rate: **a** Type of condition × type of activity effect; **b** Type of activity × behavior category. *Bars* represent standard error of the mean (SE)

The quiz game differed from all the other activities eliciting the lowest number of gaze movements away from the screen, $F(3,66) = 13.64$, $p < .001$, $\eta^2 = .38$, the highest number of facial units, $F(3,66) = 86.96$, $p < .001$, $\eta^2 = .79$, and of postural behavioral units, $F(3,66) = 19.34$, $p < .001$, $\eta^2 = .47$. The boring game elicited the lowest number of facial behavioral units. Finally, the quiz and the enemy games elicited the highest number of vocal behaviors, $F(3,66) = 22.55$, $p < .001$, $\eta^2 = .51$ (Fig. 1b).

Physiology

The features ($\mu_{X,J}$, $\sigma_{X,J}$, $\delta_{X,J}$, $\gamma_{X,J}$, $\tilde{\delta}_{X,J}$, $\tilde{\gamma}_{X,J}$) were submitted to a mixed ANOVA to test the effects of type of condition and type of activity. No significant differences emerged between the two conditions in any of the features considered. Concerning the type of activity, a significant difference between the quiz and the boring game was found on a number of features: μ, $F(3,66) = 5.13$, $p < .01$; δ, $F(3,66) = 3.04$, $p < .05$; $\tilde{\delta}$, $F(3,66) = 4.60$, $p < .01$; γ, $F(3,60) = 2.78$, $p < .05$; μ_{HR}, $F(3,60) = 2.84$, $p < .05$. Examples (i.e., EDA 1° difference, mean EDA, normalized HR) are shown in Fig. 2.

Discussion

Several authors have stressed the importance of emotion to design effective HC interactions (Picard et al. 2001), as well as the need to identify which principles that govern HH affective interactions have to be considered in order to design believable emotionally-intelligent interfaces (Cappella and Pelachaud 2001). Within this approach, our study focuses on the process of *emotional attunement* and is meant as a first attempt to structure an experimental protocol to study HC attunement. Two main hypotheses were tested. First, we expected that participants would exhibit a greater number of emotional signals towards the computer when they believed that the machine was able to understand their emotions (i.e., in the awareness condition). Second, we expected that participants would adjust (or

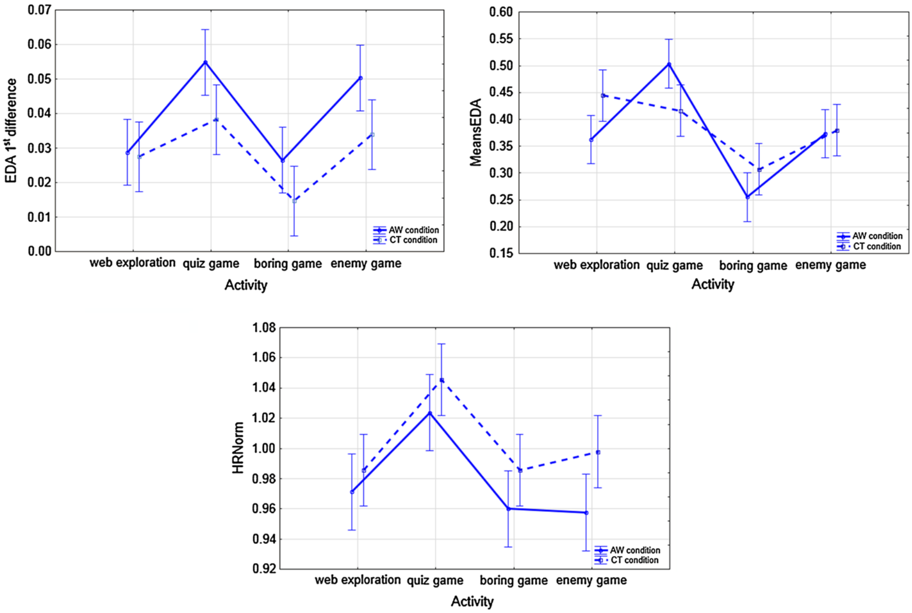

Fig. 2 Examples from the analysis of physiological signals: electrodermic activity (EDA) 1° difference, mean EDA and normalized HR. *Bars* represent standard error of the mean

attune) their nonverbal behavior to the type of activity run by the computer, showing more behavioral units during highly-interactive activities.

Do Users Exhibit a Higher Number of Emotional Signals When They Believe that the Computer is Emotionally Intelligent?

Concerning our first hypothesis, participants in the AW condition reported to believe that the computer was able to understand their emotions and to change the task significantly more than participants in the CT condition. This result seems to suggest that our manipulation was successful. The type of condition had also effect on the perceived level of involvement, even though depending on the type of activity. The AW group rated the quiz game as more involving and boring game as less involving than the CT group. No significant differences between the two groups emerged regarding either positive or negative emotional labels.

Few significant differences emerged in nonverbal behavior. The predicted tendency to exhibit more communicative signals when aware of interacting with an intentional agent was observed, but largely depending on the type of running task. In particular, participants in the AW condition showed more facial units during the quiz game (high interactivity) whereas they showed significantly less face behavior during the boring game (low interactivity) than participants in the CT condition. This result is consistent with the above reported data about the perceived level of involvement, and with the attunement process towards convergence (Giles et al. 2001). Finally, we found no significant differences in the physiological activation between the two conditions (i.e., "aware" participants were not more activated than control participants).

Overall, our hypothesis was only partially confirmed, since significance differences were few. Even though participants in the awareness condition reported to believe that the computer was able to understand their emotions, they displayed more facial behavior than participants in the control condition only when confronted with one of the two highly interactive games (the quiz), and showed less facial behavior when confronted with a low interactive, boring activity. A possible explanation for these few significances is that since an avatar talked to participants in both conditions, participants in the control condition may have attributed intentional characteristics to the computer as well (e.g., they rated that it was able to provide instructions), recognizing it as social agent (Reeves and Nass 1996). Although we do not have data to this regard, it would be interesting to compare the awareness condition to a control condition where participants interact with a standard PC (e.g., Axelrode and Hone 2005).

Do Users Adjust their Emotional Behavior to the Computer Stimuli?

Significant differences among the activities emerged in all three response domains. Concerning subjective experience, our hypotheses were confirmed. The quiz and the enemy game were judged more involving, as well as more funny, pleasant and surprising than the other two activities. Also, the boring game was rated as the most emotionally negative.

Concerning nonverbal behavior, as expected, the quiz game elicited the highest number of behavioral units and significantly differed from all the other activities with respect to all behavioral categories. Also, the boring game totalized the lowest mean scores with respect to facial movements. However, our predictions were only partially confirmed with respect to the enemy game. Although we hypothesized that both the enemy and quiz were highly interactive activities (and they were actually rated as equally involving), the enemy game elicited a significantly higher number of behavioral units than the web exploration and the boring game in the vocal behavior macro-category only. This finding may suggest that the structure of the task matters. The quiz game is characterized by a very quick and continuous alternation of questions and answers, whereas the enemy game reproduces the structure of standard arcade videogames. Emotional events are thus limited to specific points within the flow of the game.

Finally, participants were more physiologically activated when playing the quiz game than the boring game.

Overall, our results seem to indicate that the type of activity had a larger effect on participants' emotional behavior than the type of condition. Participants attuned their emotional behavior to the characteristics of the running task, being more engaged and expressing a larger number of emotional nonverbal behaviors (either facial or vocal) when they are confronted with highly interactive activities.

Limits and Future Directions

Although behavioral units were sampled through a *frame-by-frame* analysis (25 fps), leading to a total number of about 179,000 analyzed frames, a first methodological problem is the low number of participants, which may have reduced the power of analyses.

Second, future studies could employ the heart rate signal in a more extensive way, introducing the analysis of hearth rate variability for short-time series, and an additional breathing rate measure as well.

A third limit concerns our focus on the game-like activities, whereas less attention was given to the avatar, which may nonetheless play a crucial role in the attunement process. In our study, the avatar was a simple Italian speaking voice and the difference between the two conditions was limited to a number of sentences about the ability to understand emotions which were used by the avatar in the AW condition but not in the CT one. We believe that it would be interesting to further investigate the attunement process through the use of the more anthropomorphic and responsive ECAs which research has now made available.

Fourth, our four activities had different structures, and this may have led to differences in the cognitive processes demanded by each task (e.g., the quiz game and the web exploration required the reading and processing of verbal content, whereas the boring and the enemy game did not), and, possibly, on the level of involvement.

References

Axelrode, L., & Hone, K. (2005). Uncharted passions: User displays of positive affect with an adaptive affective system. *Lecture Notes in Computer Science, 3784*, 890–897.

Bailenson, J. N., Pontikakis, E. D., Mauss, I. B., Gross, J. J., Jabon, M. E., Hutcherson, C. A., et al. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human Machine Studies, 66*, 303–317.

Beebe, B., Knoblauch, S., Rustin, J., Sorter, D., Jacobs, T., & Pally, R. (2005). *Forms of intersubjectivity in infant research and adult treatment*. New York: Other Press.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior, 27*, 163–182.

Brennan, S., & Ohaeri, J. (1994). Effects of message style on users' attributions toward agents. *Proceedings of Conference Companion on Human Factors in Computing Systems* (pp. 281–282). New York, NY: ACM.

Buck, R. (1985). Prime theory: An integrated view of motivation and emotion. *Psychological Review, 92*, 389–413.

Cappella, N. J., & Pelachaud, C. (2001). Rules for responsive robots: Using human interactions to build virtual interactions. In A. L. Vangelisti, H. T. Reis, & M. A. Fitzpatrick (Eds.), *Stability and change in relationships*. New York: Cambridge University Press.

Caridakis, G., Raouzaiou, A., Karpouzis, K., & Kollias, S. (2006). Synthesizing gesture expressivity based on real sequences. In: Workshop on multimodal corpora: From multimodal behavior theories to usable models, LREC 2006 Conference, Genoa, Italy.

Cassel, J., & Bickmore, T. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction, 13*, 89–132.

Cassell, J., & Tartaro, A. (2007). Intersubjectivity in human-agent interaction. *Interaction Studies, 8*, 391–410.

Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior, 15*, 141–154.

Ciceri, R., & Balzarotti, S. (2008). From signals to emotions: Applying emotion models to HM affective interactions. In J. Or (Ed.), *Affective computing: Emotion modeling, synthesis and recognition*. Vienna, Austria: I-Tech Education and Publishing.

Ciceri, R., & Biassoni, F. (2006). Zooming on multimodality and attuning: A multilayer model for the analysis of the vocal act in conversational interactions. In G. Riva, M. T. Anguerra, B. K. Wiederhold, & F. Mantovani (Eds.), *From communication to presence* (pp. 145–165). Amsterdam: IOS Press.

Clark, H. H. (1985). Language use and language users. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 179–231). New York: Harper and RowClark.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Darves, C., & Oviatt, S. (2002). Adaptation of users' spoken dialogue patterns in a conversational interface. In: J. Hansen & B. Pellom (Eds.), Proceedings of the 7th International Conference on Spoken Language Processing. Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado, 16–20 September (pp. 561–564).Denver, CO: Casual Prod. Ltd.

de Melo, C., Carnevale, P., & Gratch, J. (2010). The influence of emotions in embodied agents on human decision-making. *Proceedings of intelligent virtual agents* (pp. 357–370). Berlin Heidelberg, Germany: Springer Verlag.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P., & Frisen, W. (1978). *Facial action coding system (FACS): A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychology Press.
Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60*, 229–240.
Frijda, N. H., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. S. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). Washington, DC: American Psychological Association.
Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.
Giles, H., Coupland, N., & Coupland, J. (1992). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge: Cambridge University Press.
Giles, H., Shepard, C. A., & Le Poire, B. A. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33–56). Chichester, UK: Wiley.
Gilleade, K.M., Dix, A., & Allanson, J. (2005). Affective videogames and modes of affective gaming: Assist me, challenge me, emote me. Paper presented at the DiGRA 2005 Conference: Changing Views–Worlds in Play.
Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. New York, NY: Anchor.
Gratch, J., & Marsella, S. (2007). The architectural role of emotion in cognitive systems. In W. Gray (Ed.), *Integrated models of cognitive systems*. Oxford: Oxford University Press.
Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
Hall, L., Woods, S., Aylett, R., Newall, L., & Paiva, A. (2005). Achieving empathic engagement through affective interaction with synthetic characters. *Lecture Notes in Computer Science, 3784*, 731–738.
Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology, 69*, 280–288.
Höök, K. (1998). Evaluating the utility and usability of an adaptive hypermedia system. *Journal of Knowledge-Based Systems, 10*, 311–319.
Höök, K. (2004). User-centred design and evaluation of affective interfaces. In Z. Ruttkay & C. Pelachaud (Eds.), *From brows to trust: Evaluating embodied conversational agents* (pp. 127–160). Dordrecht, The Netherlands: Kluwer Academic Publisher.
Izard, C. E. (1977). *Human emotions*. New York, NY: Plenum Press.
Kaiser, S., & Wehrle, T. (2001). Facial expressions as indicators of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotions: Theory, methods, research* (pp. 285–300). New York, NY: Oxford University Press.
Kappas, A., & Pecchinenda, A. (2000). Rules of disengagement: Cardiovascular changes as a function of appraisal and nine levels of difficulty of an interactive video game task. Psychophysiology, *37*, S53, Abstract.
Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13*, 505–521.
King, W. J., & Ohya, J. (1995). The representation of agents: A study of phenomena in virtual environments. *Proceedings of the IEEE International Workshop in Robot and Human Communication* (pp. 199–204). Piscataway, NJ: IEEE Press.
Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration. *Interacting with Computers, 14*, 119–140.
Kort, B., Reilly, R., & Picard, R.W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion. Paper presented at the international conference on advanced learning technologies, Madison, Wisconsin.
Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*, 730–735.
Levenson, R. W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. In P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds.), *Emotions inside out* (pp. 348–366). New York: The New York Academy of Sciences.
Lisetti, C. L. (2002). Personality, affect and emotion taxonomy for socially intelligent agents. In proceedings of the 15th international florida artificial intelligence research society conference (FLAIRS'02). Pensacola, FL: AAAI Press.
Lisetti, C. L., Nasoz, F., LeRouge, C., Ozyer, O., & Alverez, K. (2003). Intelligent affective interfaces: A patient-modelling assessment for tele-home health care. *International Journal of Human-Computer Studies, 59*, 245–255.

Koda T., & Maes, P. (1996). Agents with faces: The effects of personification of agents. Paper presented at the proceedings of the fifth IEEE international workshop on robot and human communication (RO-MAN'96).
Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*, 209–237.
Nass, C., & Lee, K. L. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Proceedings of the conference on human factors in computing systems* (pp. 329–336). New York, NY: ACM Press.
Oviatt, S., & Adams, B. (2000). Designing and evaluating conversational interfaces with animated characters. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 319–343). Cambridge, MA: MIT Press.
Oviatt, S., Darves, C., & Coulston, R. (2005). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *Transactions on Computer-Human Interaction, 11*, 300–328.
Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.
Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions Pattern Analysis and Machine Intelligence, 23*, 1185–1191.
Prendinger, H., Mayer, S., Mori, J., & Ishizuka, M. (2003). Using bio-signals to measure and reflect the impact of character-based interfaces. Paper presented at the fourth international working conference on intelligent virtual agents (IVA-03).
Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.
Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior, 31*, 87–94.
Searle, J. R. (1998). *Mind, language and society*. New York: Basic Books.
Siegman, A. W., & Feldstein, S. (1979). *Of speech and time*. Hillsdale: Erlbaum.
Stern, D. N. (1985). *The interpersonal world of the infant*. New York, NY: Basic.
Surakka, V., & Vanhala, T. (2011). Emotions in human-computer interaction. In A. Kappas & N. Krämer (Eds.), *Face-to-face communication over the internet: Emotions in a web of culture, language, and technology* (pp. 213–236). Cambridge, MA: Cambridge University Press.
Trevarthen, C. (1993). The self born in intersubjectivity: An infant communicating. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge* (pp. 121–173). New York: Cambridge University Press.
van Bakel, H. J. A., & Riksen-Walraven, J. M. (2008). Adrenocortical and behavioral attunement in parents with 1-year-old infants. *Developmental Psychobiology, 50*, 196–201.
Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology, 42*, 45–96.
Van Reekum, C. M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition Emotion, 18*, 663–688.
Zerubavel, E. (1981). *Hidden rhythms: Schedules and calendars in social life*. Chicago, IL: University of Chicago Press.