Positive influence of smile backchannels in ECAs

Elisabetta Bevacqua
CNRS - Telecom ParisTech
37/39, rue Dareau
Paris, France
bevacqua@telecom-paristech.fr

Sylwia Julia Hyniewska
CNRS - Telecom ParisTech
37/39, rue Dareau
Paris, France
hyniewska@telecom-paristech.fr

Catherine Pelachaud
CNRS - Telecom ParisTech
37/39, rue Dareau
Paris, France
pelachaud@telecom-paristech.fr

ABSTRACT

In this paper we present an evaluation study on mimicry performed by an Embodied Conversational Agent while being a listener during an interaction with a human user. Previous research has shown the importance of mimicry in human-human interaction, highlighting its relation with the level of engagement between interactants. In the present work we are interested in mimicry appearing in human-agent interactions. Through an experimental setting, we analyze humans' reactions to agent's mimicry, in particular in relation with smiles. Twelve subjects tell a story to an agent that either mimics the user's smiles, smiles randomly or does not smile at all. Results show that the agent's behavior influences the user's. Users smile more and longer when the agent performs smiling behavior. Moreover in both smiling conditions the agent is rated more positively than in the condition when it never smiles. However, no significant results have been found between the two smiling conditions.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: Evaluation/methodology

General Terms

Experimentation

Keywords

Embodied conversational agents, human-machine interaction, mimicry, smile, backchannel

1. INTRODUCTION

Being autonomous entities endowed with human-like communicative capabilities, Embodied Conversational Agents (ECAs) add a social dimension to the human-machine interaction [9, 23, 22, 18, 31]. They can talk, listen, show emotion, and so on, increasing their believability and intensifying the user's feeling of engagement with the system [29]. Research has shown that people tend to interact with computers characterized by human-like attributes as if they were real humans [30, 33]. For such a reason, a big challenge that must be faced in the design of virtual agents is the issue of credibility in the agent's behavior: during an interaction with a user, conversational agents must be able to exhibit appropriate behavior while speaking and while listening.

In the present work we are interested in the agent's listening behavior. In a conversation the interlocutor does not freeze while listening, he has to show his participation in the interaction in order to push it forward and make the speaker go on. Whenever people listen to someone, they do not assimilate passively all the words, but they actively participate in the interaction providing information about how they feel and what they think of the speaker's message. This information is transmitted through verbal and non-verbal signals, called *backchannels* [46], emitted during the speaker's turn. In accordance with the listener's behavior, the speaker can estimate how his interlocutor is reacting and can decide how to carry on the interaction: for example by interrupting the conversation if the listener is not interested or re-phrasing a sentence if the listener showed signs of incomprehension.

A particular form of backchannel is the *mimicry* of the speaker's behavior. For mimicry we mean "the behavior displayed by an individual who does what another person does" [41].

In this paper we present an evaluation study on mimicry performed by an Embodied Conversational Agent while being a listener during an interaction with a human user. Previous studies have shown the importance of mimicry in human-human interaction, highlighting its correlation with the level of engagement between interactants. In the present work we are interested in studying that mimicry can appear in human-agent interaction. Through an experimental setting, we analyze humans' reactions to the agent's mimicry. In particular we are interested in the mimicry of smile.

The next Section clarifies the concept of mimicry and how mimicry of the smile is defined. Section 3 is an overview of our agent's architecture. Then, a perceptive test we have conducted and the results we have obtained are presented.

2. STATE OF THE ART

2.1 Mimicry

Many researchers noted that during an interaction people tend to mimic several behavior. Van Baaren [41] defined mimicry as the behavior displayed by an individual who does what another person does. For example, body and leg postures, arm positions displayed by a party in a conversation are usually mimicked by the other party [27, 26, 25, 3, 36]. Facial actions mimicry has also been demonstrated. People mimic whole facial expressions of emotions [20, 19, 35], or single actions like smiling, yawning and tongue protruding [11]. Mimicry does not appear only on gestural behaviors, but also on acoustic behaviors. In [16], Giles and Powelsand saw that participants in a conversation tend to adapt to each other's accent. Studying several dyadic interviews, Webb noticed that interviewees adjusted to their interviewers' speech rate [45]. Similarly, Cappella and Panalp found that partners tend to match each other's rhythm of speech and duration of pauses [7].

Even though it was not their principal concern, all these studies showed that mimicry appears non-consciously and unintentionally. Chartrand and Bargh [10] performed a set of tests designed purposely to test the automaticity of mimicry. In their experiments, subjects participated in dyad interactions with the task of describing some photographs. In each interaction their partner was instructed to perform a particular behavior: rub her face, shake her foot, show mimicry behavior, do whatever behavior the other did not do (i.e. smiling when the other did not smile). The experiments showed that participants unintentionally tended to mimic the motor behavior performed by the other party while they worked on a task.

All these studies show that people have a natural tendency to mimic, but why does such a tendency exist? It has been shown that mimicry, when not exaggerated to the point of mocking, has several positive influences on the interaction [10, 8, 44, 11]. This type of behavior has been proven to play quite an important role during conversations. First of all, when present, it makes the conversation run more smoothly [10], and helps to regulate the conversational flow. For example, listeners often mirror speaker's postural shifts at the end of a discourse segment and this helps the exchange of speaking turn [8]. In 1964, Scheflen [36] noted that body positioning in an ongoing interaction seemed to be an indicator of liking, understanding, and the relationships between group members. Another important motivation for mimicry is its positive effect on the successfulness of the conversation that is perceived as more pleasant [44]. In [11], Chartrand et al. argue that mimicry increases empathy, liking, and rapport, binding people together. Several studies have shown that speaker's feeling of engagement increased when listeners provide backchannel signals such as nods and mimicry [17, 39, 38]. An interesting positive effect of mimicry has been shown through some tests performed by Van Baaren et al. [41] in which waitresses either verbally mimicked or did not verbally mimic their customers. Results showed that the waitresses who used the exact same words as their customer received larger tips.

2.2 Smile mimicry

The mimicry of a particular facial action is quite interesting to study: the smile. This signal has several special functions in human-human interactions. Through smile, people not only express their emotions but they provide also important information about the interaction. For example, they show the intention to start an interaction [13]; they provide backchannel signals [6] showing, for instance, their appreciation towards what the speaker is saying [5]. People who smile often are perceived more positively than people who smile less: they are seen as more attractive, friendly, warm and honest [34]. Sato and Yoshikawa [35] showed that spontaneous, externally visible facial mimicry occurs while observing dynamic facial expressions. For the mimicry of a smile (Action Unit 12 [14]) the mean latency from the onset of the dynamic expression was 817($\pm$200) ms.

In the present work we want to evaluate if smiling behavior performed by a virtual agent is perceived in a similar positive way and if interacting with an agent who smiles back is more satisfying and pleasant. In particular, we are interested in the mimicry of the smiling behavior as a form of backchannel. Several studies have shown that mimicry has positive effects on the successfulness of the conversation that is perceived as more pleasant [44]. Mimicry increases empathy, liking, and rapport, binding people together [11]. Moreover, the speaker's feeling of engagement increased when listeners provide backchannel signals such as mimicry of the speaker's behavior [17, 39, 38]. We want to study if all these positive effects of mimicry behavior are present also during user-virtual agent interactions.

2.3 Evaluation studies with ECAs

Other studies on mimicry and smiling behavior have been realized. Simons et al. [37] analyzed the behavior of users while interacting with the ECA Max to determine whether users mimic the agent's behavior. They wanted to study how agent's non-verbal behaviors affected the evaluation, feelings and the behavior of the user. Max performed eyebrow movements and self-touching gestures and participants interacted with it in two conditions: (i) occurrence of the behavior and (ii) absence of the behavior. The agent was rated as more natural, more warmhearted and more committed when presenting self-touching gestures than when it did not show any self-touching gestures. With regard to mimicry, the study did not show that participants tended to imitate any of the agent's behaviors.

Gratch et al. [18] developed the "Rapport Agent", an agent that provides solely non-verbal backchannels when listening. This agent was implemented to study the level of rapport that users feel while interacting with a virtual agent capable of providing backchannel signals. Studies performed with the "Rapport Agent" showed that the system can elicit a feeling of rapport in users through contingent behavior.

Heylen [21] looked at *how* smiles are performed during an human-human interaction and *when* these smiles appear. He determined a classification of smiles to give an insight of their different functions and to provide a clear basis on how to use smiles in virtual agent designs.

Krämer et al. [24] observed the behavior of users while interacting with the ECA Max to determine whether users mimic the agent's smile. They studied how agent's smiles affect the evaluation, feelings and the behavior of the user. Participants interacted with Max in three different conditions: (i) the agent did not smile, (ii) showed occasional smiles, (iii) smiled frequently. The agent's smile was not determined by the user's behavior. User's reaction to the agent's behavior, in particular mimicry of the agent's smile, was analyzed. Through a questionnaire, user's subjective appreciation of the agent was evaluated. The study showed that the user smiled more when the agent was smiling and that an agent that smiles less is rated more introverted.

In our evaluation, we want to study the effect of the agent's mimicry of the user's smile as a backchannel signal during a human-agent interaction. Differently from the study conducted by Krämer et al. [24], the agent performs smiles as backchannel signals and in particular conditions as a mimicry of the user's smile, while in Krämer et al. [24] the agent smiles without considering the user's behavior.

3. INTERACTIVE LISTENER ARCHITECTURE

In this section we describe our agent architecture and, in particular, the module able to generate the agent's behavior while listening to the user.

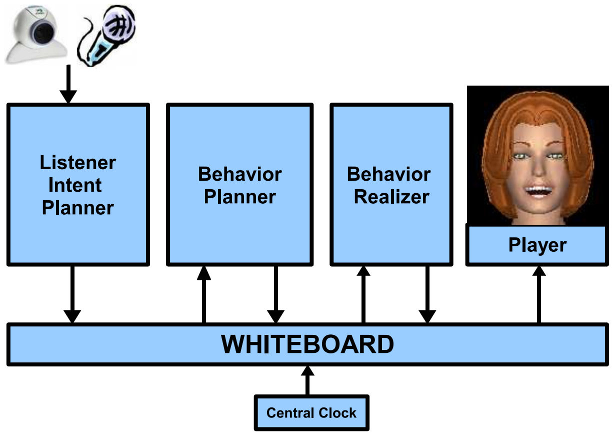

Figure 1: System architecture

3.1 System overview

Figure 1 illustrates the architecture of our system able to generate mimicry backchannel for interactive ECAs. The architecture is modular and distributed, it follows the design methodology proposed in [40] and is compatible with the standard SAIBA framework [42]. Each module exchanges information and data through a central message system. We use the concept of the *Psyclone* whiteboard [40] that allows internal modules and external software to be integrated easily. The system is designed to be used in interactive applications working in real-time.

The communicative intentions of the listener are generated by the **Listener Intent Planner** (see Section 3.2 for more details). The **Behavior Planner** module receives as input the agent's communicative intentions and generates as output a list of behavioral signals. These signals are sent to the **Behavior Realizer** that generates the animation. Finally, the animation is played in the **Player**. All components are registered to a single whiteboard to receive and send messages.

The synchronization of all modules in the distributed environment is ensured by the Central Clock which broadcasts regularly timestamps through the whiteboard.

3.2 Listener Intent Planner

The Listener Intent Planner module computes the agent's behaviors while being a listener when conversing with a user. Research has shown that there is a strong correlation between backchannel signals and the verbal and non-verbal behaviors performed by the speaker [28, 43]. From the litterature [28, 43] we have fixed some probabilistic rules to decide *when* a backchannel signal should be triggered. Our system analyzes speaker's behaviors looking for those that could prompt an agent's signal; for example, a head nod or a variation in the pitch of the user's voice will trigger a backchannel with a certain probability. Then, the system calculates *which* backchannel should be displayed. The agent can provide either signals that transmit information about its communicative functions (like agreement, liking, believing, being interested and so on) [1, 32] or signals of mimicry.

4. EVALUATION STUDY

In this evaluation we want to analyze the effect of the mimicry of the user's smile as a form of backchannel during a user-agent interaction. To this purpose, this study was designed to make subjects interact with a virtual agent in three conditions:

- *MS*: the agent provides backchannel signals and smiles only to mimic the participant when she smiles.
- *RS*: the agent provides backchannel signals smiling randomly, independently of the participant's smile.
- *NS*: the agent provides backchannel signals without smiling at all.

We hypothesize that:

- *hypothesis 1*: subjects feel more engaged in condition MS than in conditions *RS* and NS; and in condition *RS* than in NS.
- *hypothesis 2*: the interaction is seen as easier and more satisfying in condition MS than in conditions *RS* and NS; and in condition *RS* than in NS.
- *hypothesis 3*: the agent is rated more positively when it smiles during the interaction, particularly in MS condition, when it mimics the user's smile. The agent is seen as more agreeable, positive, warm, sincere and involved, as rated on a Likert scale.
- *hypothesis 4*: participants smile more in conditions MS and *RS* than in NS.
- *hypothesis 5*: participants smile longer in conditions MS and *RS* than in NS.
- *hypothesis 6*: in conditions *RS* et MS people tend to mimic the agent's smile.

Figure 2: Setting of the experiment.

4.1 Method

The setting is shown in Figure 2. Participants sat in front of the ECA displayed on a PC screen. Two video cameras recorded both the user's and the agent's behavior. Later on, videos were treated and synchronized to analyse the human-agent interaction. Twelve French speaking subjects (42% women, 58% men), mainly students, participated in this study. On average, male participants were 30.4 years old (min = 25, max = 38), whereas female subjects were 34.8 years old (min = 26, max = 50).

Each subject participated in three conditions: they were asked to read three short comic cartoon-strips (one at a time) and then tell the agent all that they remembered about the story, the characters and the drawings. Each cartoon-strip consisted of a whole short story extracted from a comic book[1].

There was no time limit for the task and participants could speak the language they felt at ease with, since there is no semantic analysis of the user's speech. The agent's behavior depends solely on the acoustic and visual behavior of the user. Generally, subjects spoke in French, however two of them preferred to use their mother language: Chinese for one and Vietnamese for the other.

They had to tell a story in each condition described above. To avoid any bias related to the order of presentation or to the nature of the presented stimuli, we defined a controlled order by associating stories and conditions as follows:

- Story A in condition MS, story B in condition RS and story C in condition NS.
- Story B in condition NS, story C in condition MS and story A in condition RS.
- Story C in condition RS, story A in condition NS and story B in condition MS.

After having told a story, subjects had to fill in a questionnaire to evaluate the agent's listening behavior during the interaction. The questionnaire, derived from that used by Gratch et al. in [18], is made of three parts:

- general information,
- perception of the agent, assessed through a set of 14 adjectives,
- a set of 20 statements evaluating the interaction.

The last two of these parts were evaluated through questions formulated in a positive and a negative manner. Subjects could rate each adjective and each statement of the questionnaire on an 8-point Likert scale (1 = disagree strongly; 8 = agree strongly). To avoid comprehension problems the questions were written in French. However, in the following Sections we will refer to each question reporting its English translation. To be sure that our questionnaire was reliable, we checked its internal consistency. Internal consistency estimates reliability by using several items to measure the same concept. We wrote two questions to measure each concept (for example, a positive formulated and a negative formulated question to evaluate the listening impression of the ECA). After collecting the responses, we run the Spearman's Rho correlation coefficient between those two questions to determine if our instrument is reliably measuring that concept. The significance level of the correlation was also checked. We expected a negative correlation between the positive and negative formulation of questions and we ruled out the positive because it would be incorrect (we looked at the one-tailed significance). The results indicate that there is a significant correlation between almost all pairs of questions (see Table 1). All correlations are negative, except one pair of adjectives where we did not obtain a significant correlation: A23 ("faked") and A30 ("spontaneous"). This may be because we did not choose the appropriate French word for the A23 adjective ("affectée"); in fact, several subjects asked for explanations about this word.

[1]"Titeuf, Tome 3, Ça épate les filles", from Zep ®Editions Glénat-1994.

Pair of questions	Spearman's Rho	
S1 - S7	-0.419	$p < 0.01$
S2 - S18	-0.261	$p < 0.05$
S3 - S8	-0.448	$p < 0.01$
S4 - S13	-0.609	$p < 0.01$
S5 - S15	-0.765	$p < 0.01$
S6 - S10	-0.647	$p < 0.01$
S9 - S14	-0.864	$p < 0.01$
S11 - S19	-0.683	$p < 0.01$
S12 - S17	-0.415	$p < 0.01$
S16 - S20	-0.659	$p < 0.01$
A21 - A28	-0.786	$p < 0.01$
A22 - A33	-0.699	$p < 0.01$
A23 - A30	0.170	$p = 0.09$
A24 - A27	-0.725	$p < 0.01$
A25 - A34	-0.741	$p < 0.01$
A26 - A31	-0.757	$p < 0.01$
A29 - A32	-0.599	$p < 0.01$

Table 1: Results of the Spearman's Rho correlation coefficient.

4.2 Participants and equipment

Participants entered the laboratory one at a time and were informed that we were performing a study to evaluate a virtual agent in the role of the listener. The experimenter explained each subject what a virtual agent is and what she had to do. After having read and signed the consent form, the participant was asked to sit in front of a screen and was introduced to the virtual agent and to the equipment used in the experiment. The subject read the first short comic-strip and, to make her at ease, she was left alone to tell the story to the agent. Before proceeding with the other comic-strips the participant was asked to fill in the questionnaire.

During the interaction the agent provides only positive backchannel signals to show it is listening and to incite the participant to go on. Possible backchannels are: raise of the eyebrows, head nod, smile and all their combinations [4]. To add some variability in the agent's behavior three types of head nod (slow, medium and fast) and three types of smile (small, medium and wide) were displayed randomly: that is, the type of smile displayed by the agent was not linked to the subject's behavior. The effect on the user of these three different types of smile has not been evaluated.

Even if our system is able to generate backchannels signals according to the user's non-verbal behavior, in order to work fine it needs reliable video and audio information. Since we do not have at disposition a reliable and robust application to recognize user's smiling behavior, a Wizard of Oz setting was used. In another room the experimenter drove the system to provide signals of smile. Looking at the participant through a webcam, the experimenter provided a backchannel each time a pause in the user's voice occurred, or when a pitch change was perceived (like at the end of an exclamation or a question) or when the user was smiling (any type of smiles was considered). Two buttons on a graphic interface allowed the experimenter to select either a backchannel with a smile or a backchannel without a smile. Backchannels with a smile were selected to mimic user's smiles in the MS condition or to provide random smiles in the RS condition.

4.3 Results

All participants (N=12) gave responses to the statements in each condition. The Friedman-test was used for this repeated-measures design. Results show that there is an effect of the condition only for three statements: (i) "warm" ($\chi^2 = 6.5$, $df = 2$, $p = 0.039$), "positive" ($\chi^2 = 6.5$, $df = 2$, $p = 0.039$) and "I think that the agent wasn't really listening to me" ($\chi^2 = 6.07$, $df = 2$, $p = 0.048$).

We also compared the answers to each question pair-wise, between each couple of conditions. We have a within participants design, as the same participants are being measured in all three conditions. We used the Wilcoxon to compare pair-wise the answer to each question. The Wilcoxon test showed significant differences for some of the questions. Subjects felt less engaged in condition NS than in condition MS ($p < 0.05$). They judged the agent less "positive" ($p < 0.05$) and less "warm" ($p < 0.05$) in condition NS than in condition RS. A difference appears also between conditions NS and MS ($p < 0.05$). The agent appeared more interested in the condition RS, where it smiles without mimicry, than in condition NS ($p < 0.05$). The interaction has been judged more frustrating in condition NS than in MS ($p < 0.05$). Finally, participants felt more at ease ($p < 0.05$) and more listened to ($p < 0.05$) while telling the story to the agent in condition MS than RS.

All the smiles performed by both the agent and the user were annotated in the three conditions. Facial actions considered by the WOZ as smiles were verified to be either AU 12 (lip corner puller) or AU 14 (dimpler) [15]. The timing of the process involved in the mimicry condition was evaluated. The reaction from the WOZ was between 800-1000 ms. from the apex of the user's smile. The time of the animation synthesis was between 500-700 ms. The reliability of annotation for the frequency of smiles was assessed for 17% (6 videos, 2 per condition) of the data, realized by a second coder who was FACS (Facial Action Coding System) certified. Agreement was assessed with Cohen's kappa [12], which evaluates inter-observer agreement after correcting for level of agreement expected by chance. The mean kappa across conditions was 0.93.

We calculate the frequency of the user's smiles (FS) as the total number of smiles divided by the duration of the interaction in seconds. Table 2 shows the corresponding descriptive statistics.

The mean frequency of smiles per second is 0.060 in condition MS (standard deviation 0.042), 0.042 in RS (standard deviation 0.034) and 0.028 in NS (standard deviation 0.029). The data have not a normal distribution. Looking at the box plot diagram (Figure 3), one could argue that the mean frequency of smile is higher in condition MS than in RS and NS and it is higher in condition RS than in NS. We

	FS_RS	FS_MS	FS_NS
mean	0.04198	0.0604	0.02864
standard error	0.0100	0.0121	0.0084
standard deviation	0.0349	0.0420	0.0293
minimum	0	0	0
maximum	0.1	0,12	0,08
level of confidence	0.0221	0.0266	0.0186

Table 2: Descriptive statistics of the frequency of the user's smiles (N=12).

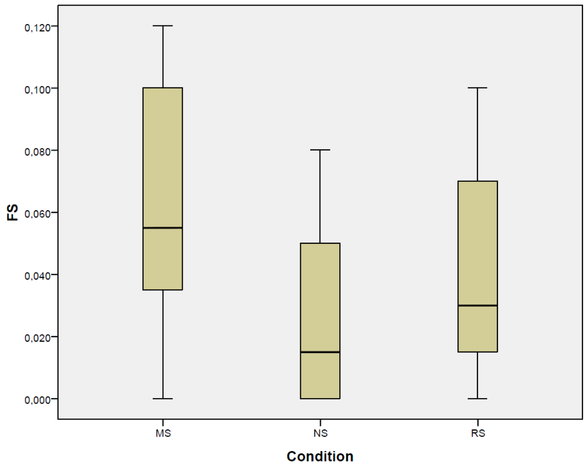

Figure 3: Box plot diagram of the mean frequency of smiles.

checked if such differences were significant. Through Friedman test we obtained a significant difference between the three conditions ($p < 0.05$). Wilcoxon test showed a difference between the conditions MS and NS ($p < 0.05$). The difference between the conditions RS and NS was on the limit of significance ($p = 0.052$). No significant difference was found between the conditions MS and RS ($p = 0.117$).

We also calculated the mean duration of smiles (MDS) as the total duration of smiles divided by the number of smiles. Table 3 shows the corresponding descriptive statistics.

	MDS_RS	MDS_MS	MDS_NS
mean	1.42	1.5841	0.8941
standard error	0.1469	0.2789	0.2122
standard deviation	0.5090	0.9661	0.7351
minimum	0	0	0
maximum	2	3	2,04
level of confidence	0.3234	0.6138	0.4671

Table 3: Descriptive statistics of the mean duration of user's smiles (N=12).

The mean duration of smiles per second is 1.58 in condition MS (standard deviation 0.966), 1.42 in RS (standard deviation 0.509) and 0.89 in NS (standard deviation 0.735). The data have not a normal distribution.

We applied the Wilcoxon test and we looked at (1-tailed) Exact sign. We obtained a significant difference between the conditions RS and NS ($p < 0.05$) and the conditions MS and NS ($p < 0.05$). No significant difference was found

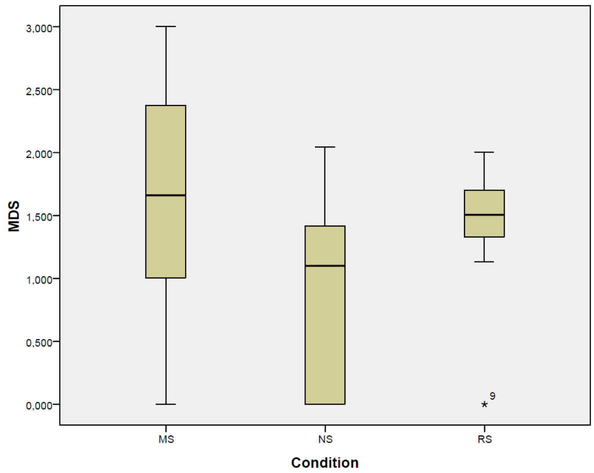

Figure 4: Box plot diagram of the mean duration of smiles.

between the conditions MS and RS ($p > 0.05$).

Finally we computed the number of smiles performed by the user as a mimicry of the agent's smile (SM). We did not consider participants' smiles in condition NS as being a mimic signal since the agent never smiled in this condition. Since we found that the user's mimicry depends on the number of times the agent smiles, we calculated the ratio of mimicked smiles (RMS) as the user's mimicked smiles divided by the number of the agent's smiles.

The descriptive statistics are reported in Table 4.

	RMS_RS	**RMS_MS**	**RMS_NS**
mean	0.4583	0.3847	0
standard error	0.0914	0.0733	0
standard deviation	0.3168	0.2541	0
minimum	0	0	0
maximum	1	0.8571	0
level of confidence	1.3226	1.5903	0

Table 4: Descriptive statistics of the ratio of user's mimicked smiles (N=12).

In condition MS the mean ratio of user's mimicked smiles is 0.38 whereas in condition RS it is 0.45. We did not obtain any significant difference between the two conditions.

4.3.1 Results analysis

To test our first hypothesis, that subjects feel more engaged in condition MS than in the conditions RS and NS (and in condition RS than in NS), we looked at the statements "I felt I was able to engage the agent with my story" and "I felt I wasn't able to engage the agent with my story". For the first statement we did not find significant differences between the conditions. On the other hand, we found a significant difference between how subjects judged the second statement in the condition MS and in NS. People felt less engaged when the agent did not display any smile and they felt significantly more engaged when the agent mimicked their smiles as a form of backchannel. Such a result sustains in part our first hypothesis.

With regard to the second hypothesis, that the interaction is seen as easier and more satisfying in condition MS than in conditions RS and NS, and in condition RS than in NS, again we obtain just partial evidence. Participants judge the interaction significantly more frustrating in condition NS, when the agent never smiles, than in the condition MS, but we have no significant difference either between the conditions MS and RS or between the conditions RS and NS.

The agent has been seen as warmer and more positive in both conditions MS and RS than in condition NS. Participants have a more positive impression of the agent when it smiles both to mimic the user's smile and according to a random pattern. This sustains our third hypothesis, that the agent is rated more positively when it smiles during the interaction. No significant differences where found between the conditions MS and RS; we think that, in general, people are sensitive to the agent's smile, but they did not find much difference between smiles derived from the mimicry of their own smile and "random" smiles, who are still appropriate as backchannel signals, even without relying on mimicry.

Analyzing the videos we noticed that people smile more when the agent displays a smiling behavior, whether in condition MS or in condition RS. In particular, we found a significant difference between the frequency of user's smiles in condition MS and in condition NS, sustaining our fourth hypothesis, that participants smile more in condition MS than in NS. We did not obtain significant differences between the condition RS and NS.

Moreover, we found that people tend to smile longer when the agent shows some smiling behavior. The duration of the agent's smile has not been analyzed, since it was always the same during all interactions. Results show that there are significant differences between the mean duration of smiles in condition MS and NS and between the mean duration of smiles in condition RS and NS. No significant results were found for condition MS and RS. That sustains our fifth hypothesis, that is participants smile longer in conditions MS and RS than in NS. As regard to the sixth hypothesis, that in conditions RS and MS people tend to mimic the agent's smile, we did not obtain significant statistic results. We observed also that, in condition MS, subjects tend to make their smile more intense when they notice that the agent smiles back. In the future intensity could be coded with the FACS criteria for the targeted action units and for a greater number of participants.

4.4 Discussion

We observed the latency of the WOZ's action (i.e. sending a backchannel command to the agent) to be always below 1000 ms in the participants' interactions. The computation time of the agent animation is between 500 et 700 ms. So, all in all, the time delay between a user's smile and an agent's smile is below 1700 ms. While this timing is larger than what is observed in spontaneous human facial mimicry [35], we believe that the agent's smile's contingency is sufficient to have an effect on the interaction [2]. Our study reinforces the results found by Gratch et al. [17] through the studies with the "Rapport Agent". Like them we saw that interacting with a virtual agent able to perform contingent backchannel signals increases the feeling of rapport and engagement in users. Our study shows, moreover, that the agent's **smile** backchannel has an influence on the user's behavior: users feel more engaged when the ECA smiles back and the agent is rated more positively.

In their study Kramer and colleagues [24] investigated similarly the effect of smiling behavior on the perception of the agent. While their results stay at a non-significant level, we can confirm that in our study there was a clear increase in the positivity of the rating when the agent smiled. We did not find a significant difference between the rating of an agent that shows random smile backchannels and one that shows mimicked smile backchannels as we hypothesized. Maybe, to make a virtual agent be perceived more positively, it is enough that the ECA performs smile backchannels even if the user does not smile. As other studies have shown that mimicry influences engagement [17, 39, 38], it could be interesting to study what modulates this effect. The effect could be stronger, for example, with an increase of the duration of the interactions, as our interaction time was on average not longer than 2 minutes.

Through our test we saw also that participants tend to mimic the agent's smile and, even if we did not obtain significant statistic results to differentiate between the random smile and the mimicked smile conditions, the observation of the videos allowed us to gather some interesting information. First of all, we noticed that in both smiling conditions people often smiled back and when they did not respond to the smile usually they were not looking at the agent, so they could not see the agent displaying a smile. For example, even the two subjects who never smiled in the condition in which the agent did not smile performed at least one short smile as a response to the agent's smile in the condition in which the agent performed random smiles. No significant results were found when comparing between the two smiling conditions and we think that in general users are not necessarily more sensitive to the agent's mimicked smiles. It is the agent's smiling behavior that has an impact on the user's perception of the agent, independently of the fact that the agent's smile derives from mimicry or not.

These results show that ECAs developers could, before all, take into account the agent's smiling behavior per se, and not particularly a contingent one, since it seems to influence the quality of the user-agent interaction.

5. CONCLUSION

We have presented an evaluation study conducted on backchanneling including smiles, mimicked smiles and other non-smiling backchannels. These were performed by an Embodied Conversational Agent in the role of the listener during an interaction with a human user. Results show that the agent's behavior influences positively the user's. Users smiled more and longer when the agent performed some smiling behavior. Moreover in both smiling conditions the agent was rated more positively than in the condition in which it never smiled.

6. ACKNOWLEDGMENTS

This work has been supported by the STREP SEMAINE project IST-211486 (http://www.semaineproject.eu).

7. REFERENCES

[1] J. Allwood, J. Nivre, and E. Ahlsén. On the semantics and pragmatics of linguistic feedback. *Semantics*, 9(1), 1993.

[2] J. Bailenson, A. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 13:428–441, 2004.

[3] F. Bernieri and R. Rosenthal. Interpersonal coordination: Behavior matching and interactional synchrony. In R. Feldman and B. Rime, editors, *Fundamentals of nonverbal behavior. Studies in emotion and social interaction*, pages 401–432. University Press, New York: Cambridge, 1991.

[4] E. Bevacqua, D. Heylen, M. Tellier, and C. Pelachaud. Facial feedback signals for ECAs. In *AISB'07 Annual convention, workshop "Mindful Environments"*, pages 147–153, Newcastle upon Tyne, UK, April 2007.

[5] E. Bevacqua, M. Mancini, and C. Pelachaud. A listening agent exhibiting variable behaviour. In H. Prendinger, J. C. Lester, and M. Ishizuka, editors, *Proceedings of 8th International Conference on Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, pages 262–269, Tokyo, Japan, 2008. Springer.

[6] L. Brunner. Smiles can be back channels. *Journal of Personality and Social Psychology*, 37(5):728–734, 1979.

[7] J. Cappella and S. Planalp. Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research*, 7(2):117–132, 1981.

[8] J. Cassell, Y. Nakano, T. Bickmore, C. Sidner, and C. Rich. Non-verbal cues for discourse structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 114–123. Association for Computational Linguistics Morristown, NJ, USA, 2001.

[9] J. Cassell, J. Sullivan, P. Prevost, and E. Churchill. Embodied conversational characters. 2000.

[10] T. Chartrand and J. Bargh. The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Personality and Social Psychology*, 76:893–910, 1999.

[11] T. Chartrand, W. Maddux, and J. Lakin. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The new unconscious*, pages 334–361, 2005.

[12] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37, 1960.

[13] I. Eibl-Eibesfeld. *Die Biologie des menschlichen Verhaltens: Grundriss der Humanethologie*. Piper, München, 1997.

[14] P. Ekman and W. Friesen. *Manual for the Facial Action Coding System*. Palo Alto: Consulting Psychologists Press, 1977.

[15] P. Ekman, W. Friesen, and J. Hager. The facial action coding system. In *Research Nexus eBook*, 2002.

[16] H. Giles and P. Powesland. *Speech Style and Social Evaluation*. Not Avail, 1975.

[17] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. van der Werf, and L. Morency. Virtual Rapport. In J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier, editors, *Intelligent Virtual Agents (IVA'06). LNCS (LNAI)*, volume 4133, Heidelberg, 2006. Springer.

[18] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In C. P. et al., editor, *7th International Conference on Intelligent Virtual Agents*, Paris, France, 2007.

[19] U. Hess and S. Blairy. Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40(2):129–141, 2001.

[20] U. Hess, P. Philippot, and S. Blairy. Mimicry: Facts and fiction. *The social context of nonverbal behavior*, pages 213–241, 1999.

[21] D. Heylen. Talking head says 'cheese' - humor as an impetus for embodied conversational agent research. In *Workshop at CHI 2003*, Fort Lauderdale, Florida, USA, April 2003.

[22] D. Heylen. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics (IJHR)*, 3(3), 2006.

[23] S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.

[24] N. C. Krämer, N. Sommer, S. Kopp, and C. Becker-Asano. Smile and the World Will Smile With You: The Effects of a Virtual Agentt's Smile on Users' Evaluation and Smiling Behavior. In *Paper presented at the annual meeting of the International Communication Association*, Marriott, Chicago, 2009.

[25] M. La France and M. Broadbent. Group rapport: Posture sharing as a nonverbal indicator. *Group and Organization Studies*, 1:328–333, 1976.

[26] M. LaFrance. Posture mirroring and rapport. *Interaction rhythms: Periodicity in communicative behavior*, pages 279–298, 1982.

[27] M. LaFrance and W. Ickes. Posture mirroring and interactional involvement: Sex and sex typing effects. *Nonverbal Behavior*, 5(3):139–154, 1981.

[28] R. M. Maatman, J. Gratch, and S. Marsella. Natural behavior of a listening agent. In *5th International Conference on Interactive Virtual Agents*. Kos, Greece, 2005.

[29] S. V. Mulken and E. André. The persona effect: How substantial is it. pages 53–66. Springer, 1998.

[30] C. I. Nass, Y. Moon, J. Morkes, E. Y. Kim, and B. J. Fogg. Computers are social actors: a review of current research. pages 137–162, 1997.

[31] C. Pelachaud. Multimodal expressive embodied conversational agents. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 683–689, New York, NY, USA, 2005. ACM.

[32] I. Poggi. *Mind, hands, face and body. A goal and belief view of multimodal communication.* Berlin: J. Weidler, 2007.

[33] B. Reeves and C. Nass. The media equation: How people treat computers, television and new media like real people and places. 1996.

[34] H. Reis, I. Wilson, C. Monestere, S. Bernstein, K. Clark, and E. Seidl. What is smiling is beautiful and good. *European Journal of Social Psychology*, 20(3):259, 1990.

[35] W. Sato and S. Yoshikawa. Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, 104(1):1–18, 2007.

[36] A. E. Scheflen. The significance of posture in communication systems. *Psychiatry*, 27:316–331, 1964.

[37] N. Simons, N. Kraemer, and S. Kopp. The effects of an embodied conversationational agent's subtle nonverbal behavior on user's evaluation and behavioral mimicry. In *57th Annual Conference of the International Communication Association (ICA 07)*, San Francisco, CA, 2007.

[38] J. Smith. *GrandChair: Conversational Collection of Family Stories.* Media Lab, MIT, Cambridge, MA, 2000.

[39] D. Tatar. *Social and personal consequences of a preoccupied listener.* Ph.D. Dissertation, Department of Psychology. Stanford University, 1998.

[40] K. R. Thórisson, T. List, C. Pennock, and J. Dipirro. Whiteboards: Scheduling blackboards for semantic routing of messages & streams. In *AAAI-05 Workshop on Modular Construction of Human-Like Intelligence*, pages 8–15, 2005.

[41] R. B. Van baaren. Mimicry: a social perspective, 2003. webdoc.ubn.kun.nl/mono/b/baaren_r_van/mimi.pdf.

[42] H. H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf. The Behavior Markup Language: Recent developments and challenges. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, editors, *Proceedings of 7th International Conference on Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, pages 99–111, Paris, France, 2007. Springer.

[43] N. Ward and W. Tsukahara. Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics*, 23:1177–1207, 2000.

[44] R. M. Warner, D. Malloy, K. Schneider, R. Knoth, and B. Wilder. Rhythmic organization of social interaction and observer ratings of positive affect and involvement. *Journal of Nonverbal Behavior*, 11(2):57–74, 1987.

[45] J. T. Webb. Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. *Studies in dyadic communication*, pages 115–133, 1972.

[46] V. Yngve. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–577. 1970.