Evaluation of a virtual listener's smiling behavior

Elisabetta Bevacqua, Sylwia Julia Hyniewska and Catherine Pelachaud
CNRS - Telecom ParisTech
{bevacqua,hyniewska,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. Through an experimental setting, we analyze humans' reactions to agent's mimicry, in particular in relation with smiles. Results show that the agent's behavior influences the user's.

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

1 Introduction

Embodied Conversational Agents (ECAs) add a social dimension to the human-machine interaction [4, 7]. They can talk, listen, show emotion, and so on, increasing their believability and intensifying the user's feeling of engagement with the system. For such a reason, 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. Whenever people listen to someone, 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* [9], emitted during the speaker's turn. A particular form of backchannel is the *mimicry* of the speaker's behavior.

We present an evaluation study on mimicry performed by an ECA while being a listener during an interaction with a user. Through an experimental setting, we analyze humans' reactions to the agent's mimicry of smile. The next Section clarifies the concept of mimicry. Then, a perceptive test we have conducted and the results we have obtained are presented.

2 State of the art

Many researchers noted that during an interaction people tend to mimic behavior. Mimicry appears on gestural behaviors, like facial expressions or body and leg postures and on acoustic behaviors [3]. It has been shown that mimicry has several positive influences on the interaction, like, for example, making the conversation more pleasant. Chartrand et al. [3] argue that mimicry increases empathy and rapport, liking and binding people together.

In the present work we study if all these positive effects of mimicry behavior are present also during user-virtual agent interactions. In particular, we are interested in the mimicry of the smiling behavior as a form of backchannel. The mimicry of the smile has several special functions in human-human interactions. For example, through smile, people show the intention to start an interaction or they provide backchannel signals [2]. Other studies on mimicry and smiling behavior have been realized. Krämer et al. [5] studied how agent's smiles affect the evaluation, feelings and the behavior of the user while interacting with the ECA Max. 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 backchan-

nel. Differently from the study conducted by Krämer et al. [5], the agent performs smiles as backchannel signals and in particular conditions as a mimicry of the user's smile, while in Krämer et al. [5] the agent smiles without considering the user's behavior.

3 Evaluation study

The ECA system used to perform this study is described in [6]. 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:

- *hp1*: subjects feel more engaged in condition MS than in conditions RS and NS; and in condition RS than in NS.
- *hp2*: 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.
- *hp3*: the agent is rated more agreeable, positive, warm, sincere and involved when it smiles during the interaction.
- *hp4*: participants smile more in conditions MS and RS than in NS.
- *hp5*: participants smile longer in conditions MS and RS than in NS.
- *hp6*: in conditions RS et MS people tend to mimic the agent's smile.

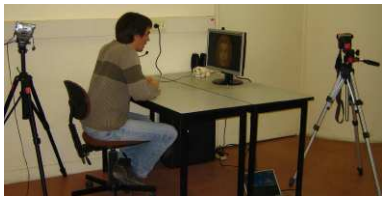

Figure 1: Setting of the experiment.

3.1 Method and participants

Figure 1 shows the setting of the evaluation. 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, whereas female subjects were 34.8. Subjects 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. They had to tell a story in each condition described above. There was no time limit for the task. After having told a story, subjects had to fill in a questionnaire (derived from that used by Gratch et al. in [4]) to evaluate the agent's listening behavior during the interaction. Participants could rate each statement of the questionnaire on an 8-point Likert scale (1 = disagree strongly; 8 = agree strongly). To be sure that our questionnaire was reliable, we wrote two questions to measure each concept (for example, a positive formulated and a negative formulated question).

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 [1]. To generate backchannels signals according to the user's non-verbal behavior, our system needs reliable video and audio information. Since we do not have at disposition a reliable and robust application a Wizard of Oz setting is used. In another room the experimenter drove the system to provide signals of smile. 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). Backchannels containing a smile were selected to mimic user's smiles in the MS condition or to provide random smiles in the RS condition.

3.2 Results

At first, we checked that our questionnaire was reliable. After collecting the responses, we run the Spearman's Rho correlation coefficient between each pair of questions. The significance

level of the correlation was checked. The results indicate that there is a significant correlation between almost all pairs of questions.

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: "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 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. These results sustain our first three hypotheses.

All the smiles performed by both the agent and the user were annotated in the three conditions. We calculate the frequency of the user's smiles as the total number of smiles divided by the duration of the interaction in seconds. 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, the mean kappa across conditions was 0.93. The mean frequency of smiles per second is 0.06 in condition MS (sd 0.042), 0.042 in RS (sd 0.034) and 0.028 in NS (sd 0.029). 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 as the total duration of smiles divided by the number of smiles. The mean duration of smiles per second is 1.58 in condition MS (sd 0.966), 1.42 in RS (sd 0.509) and 0.89 in NS (sd 0.735). 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 between the conditions MS and RS ($p > 0.05$). That sustains our fifth hypothesis.

Finally we computed the number of smiles performed by the user as a mimicry of the agent's smile. 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 as the user's mimicked smiles divided by the number of the agent's smiles. 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.

3.3 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 [8], we believe that the contingency of the agent's smile is sufficient to have an effect on the interaction.

In their study Kramer and colleagues [5] 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. However, 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. We

think that, to make a virtual agent be perceived more positively, it is important that it performs smile backchannels independently of the users' smile.

Through our test we saw also that participants smile longer and more often when the agent smiles. Moreover they 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. 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.

4 Conclusion

We have presented an evaluation study conducted on backchanneling including smiles, mimicked smiles and other non-smiling backchannels. These were performed by an ECA 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 longer and more often 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.

Acknowledgments

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

References

[1] 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.

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

[3] 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.

[4] 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.

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

[6] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta: an interactive expressive eca system. In *AAMAS'09*, Budapest, Hungary, 2009.

[7] 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.

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

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