PaolaChat: A Virtual Agent with Naturalistic Breathing

David Novick(✉), Mahdokht Afravi(✉), and Adriana Camacho(✉)

The University of Texas at El Paso, El Paso, TX 79968, USA
novick@utep.edu, mmafravi@gmail.com, caro4874@gmail.com

Abstract. For embodied conversational agents (ECAs) the relationship between gesture and rapport is an open question. To enable us to learn whether adding breathing behaviors to an agent similar to SimSensei would lead users interacting to perceive the agent as more natural, we built an application, called Paola Chat, in which the ECA could display naturalistic breathing animations. Our study had two phases. In the first phase, we determined the most natural amplitude for the agent's breathing. In the second phase, we assessed the effect of breathing on the users' perceptions of rapport and naturalness. The study had a within-subjects design, with breathing/not-breathing as the independent variable. Despite our expectation that increased naturalness from breathing would lead users to report greater rapport in the breathing condition than in the not-breathing condition, the study's results suggest that the animation of breathing appears to neither increase nor decrease these perceptions.

Keywords: Embodied conversational agents · Human-agent dialog
Dialog system

1 Introduction

The relationship between embodied conversational agents' (ECAs) [1] gestures (see e.g., [2]) and rapport (see e.g., [3, 4]) is a currently active research question. While some studies have reported the effect of users' perception of extraversion on gesture amplitude [5], other studies reported that gesture amplitude may not affect users' perception of rapport [6]. This disparity suggests that the naturalness of ECAs' gestures may be a significant factor in shaping users' perceptions. Indeed, not all ECAs have the same level of naturalness of behavior.

The building of human-ECA rapport is increasingly important as ECAs take on more meaningful roles, including serving as a means of diagnosing PTSD [7]. SimSensei, the PTSD-diagnosis agent, while having excellent animation of facial features, appeared to have a static torso, which might give the impression that she is holding her breath or not breathing between her utterances. Therefore, we sought to answer the question of whether adding naturalness, in this case for breathing, would lead to higher perceptions of rapport.

To this end, we built an application that would enable us to learn whether adding breathing behaviors to a similar agent would lead humans to perceive the agent as more

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 351–360, 2018.
https://doi.org/10.1007/978-3-319-91581-4_26

natural and to develop a higher level of rapport with the agent. The application we developed, called Paola Chat, featured an ECA named Paola that resembled SimSensei but was able to display naturalistic breathing animations. The animations were based on a simple model of respiration and an empirical study of the perceived naturalism of breathing amplitude. The application enabled study of whether users' perceptions of the ECA's naturalness would increase based on the varying frequency and amplitude of the ECA's breathing during the conversation.

2 Implementation

This study comprised two phases: a preliminary phase in which we sought to avoid the possible effects non-standard amplitude could have on perceptions of naturalness (cf., [6]) and a second phase in which we assessed the effect of breathing on users' perceptions of rapport.

2.1 Phase 1: Naturalistic Amplitude

In the first phase, we sought to find the most natural amplitude of gestures for perceptions of naturalness (cf., [6].) To provide the Paola Chat agent with amplitude of breathing that would be as natural as possible, we conducted an empirical evaluation of human perceptions of the agent with different amplitudes for the breathing animation. We prepared five brief animations of the ECA's breathing, ranging from static (not breathing) to exaggerated breathing. We sought to have the agent show breathing with amplitude large enough to be salient but not so large as to appear unnatural or distracting. Each participant viewed the five representative animations as an introduction to the task. Each participant then saw and rated 30 animations for naturalness, presented in random order, using a five-point Likert scale.

Unsurprisingly, we found that that the animation rated the most natural was the medium-amplitude animation, which had a 50% amplitude from the extreme. Figure 1 compares Paola's breathing at the last point of inhale (just before the transition to exhale) with Paola's breathing at the last point of exhale (just before the transition to inhale) in this medium amplitude; the differences apparent in this static figure are subtle.

2.2 Phase 2: Effect on Perception of Rapport

With the amplitude determined, in the study's second phase we assessed the effect of breathing on the users' perceptions of rapport. To this end, we implemented Paola Chat, which is a fully automated conversational agent rather than a Wizard-of-Oz system. We designed Paola to resemble SimSensei as much as possible. Paola, displayed as a life-sized person projected on the wall in UTEP's immersion laboratory, was seated in a large chair, with her legs crossed and her hands resting on her lap, resting on the chair's armrests, or making gestures while speaking (see Fig. 2). Users' interactions with Paola consisted of two back-to-back conversations on the topics of vacations and movies. For example, in the movie conversation, Paola asked "Have you seen any movies lately?"

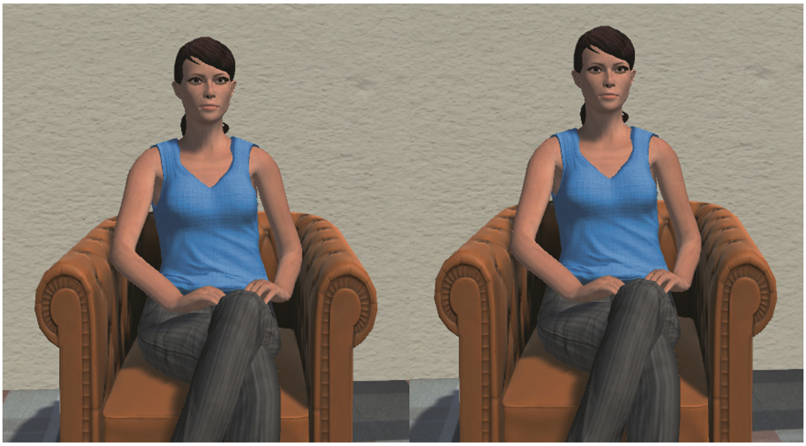

Fig. 1. Paola's inhale before transition (left) to exhale (right).

Paola would then interpret the user's response using keyword recognition (i.e., if the user answered that he or she had, Paola would ask for details about the movie, but if the user had responded in the negative, Paola would segue to another question about the user's favorite movie).

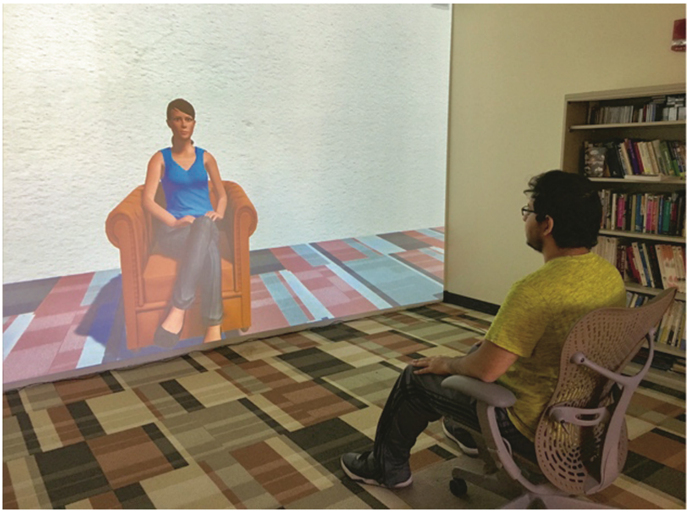

Fig. 2. A person interacts with the Paola Chat agent.

Developing this application required accepting a wide range of dialogue input and generating relevant responses. Paola Chat was developed with the UTEP AGENT system [8], which is capable of accepting a wide range of utterances, called wildcards, irrespective of topic or word choice. A major difficulty was that the study needed to have Paola's utterances generate extended responses from the users so that they could observe

Paola both while she was talking and while she was listening. But, consequently, the system must be able to generate responses that are generic enough to keep the conversation from seeming one-sided or disconnected. Some questions were posed as “yes” or “no” response questions, in which case the dialogue tree would converge back to a certain point to naturalize the flow of utterances generated by Paola.

Breathing Model. The agent needed a function to control the breathing with respect to three constraints. First, the agent should not be speaking while displaying an inhaling animation. Second, the agent should appear to take breaths of natural length between utterances. Third, the agent’s frequency (in addition to amplitude) of breathing should be perceived as natural.

Breathing includes the states of inhaling or exhaling, and their transitions [8]. Our system required the development of a model that could represent these states smoothly, as well as having amplitude, oscillation, and frequency. In our model, the breathing state depends on the amplitude and oscillation. The amplitude represents the *y*-value on a graph and visually represents how much the ECA’s torso would expand. The oscillation represents the *x*-value on a graph in radians. The oscillating state (i.e., the wave) moves depending on the frame rate of the animation and the frequency per frame. The frequency was set to a fixed value per frame, adjusted for frame drops because Unity sometimes skips frames. During the interaction, the oscillations varied between 100 and 0, as an effect of the sine wave function:

$$\text{Breathing State} = \text{Amplitude} + (\text{Amplitude} \times \sin(\pi \times \text{Frequency}))$$

The breathing oscillating function ran in a cycle in which the agent would either inhale (breathing state = 0) or exhale (breathing state = 100). The cycle was interrupted only when the agent was about to speak, to portray an in-breath before speech. Figure 3 shows the breathing oscillation function: the x-axis is the oscillation of sin(πx), where x is the frequency, and the y-axis is the changes in the wave function for the values of the updating breathing state, with 0 the lowest point of exhale and 100 as the highest point of inhale. Figure 4 depicts the transitions between the states in the breathing oscillating function.

Application Dialog. We deployed Paola Chat’s breathing model in a pair of conversations about vacations and movies. The length of the conversations ranged between five to seven minutes, depending on user responses.

The UTEP AGENT system [9], in which Paola Chat was implemented, interfaces with the Unity game engine to automate features during the interaction, i.e., generating dialogue, handling breathing and other gesture animations, cycle through the states of breathing, as well as recognizing user input during conversation .

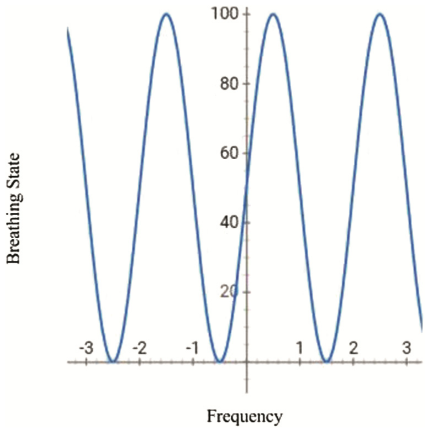

Fig. 3. Breathing oscillation function. The x-axis is the oscillation of sin(πx), where x is the frequency, and the y-axis is the change in the wave function for the values of the updating breathing state, with 0 the lowest point of exhale and 100 as the highest point of inhale.

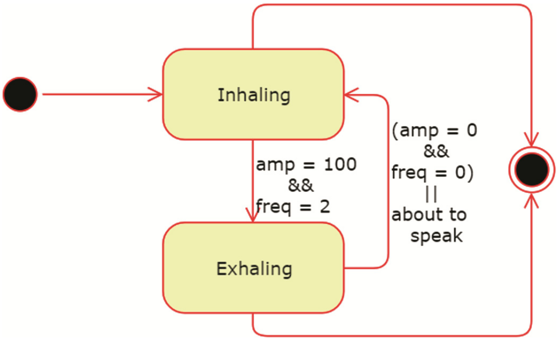

Fig. 4. The breathing cycles represented as a state diagram.

In the first conversation, Paola greeted the participant and began conversing on the topic of either vacations or movies; the order of the topics alternated as a part of the experimental design. Paola would occasionally ask questions where the participant's utterance would be either treated as a wildcard (where the content did not matter) or, based on keyword recognition, would trigger an appropriate response.

Table 1 shows examples of responses to questions asked by Paola and follow-up questions Paola asked during the interactions. For example, Paola would ask "*Have you seen any movies lately?*" If the participant responded that he or she had, then Paola would ask for more details about that movie. If, however, the participant responded no, then Paola would instead say "*It's okay! Tell me about your favorite movie, then. What is your favorite?*"

Table 1. Responses and follow-up questions.

Paola's utterance	*Example user response*	*Paola's follow-up*
Have you seen any movies lately?	Yes, I have	Oh, interesting. What was that movie about, was it good?
Have you seen any movies lately?	No, I haven't watched too many movies lately	Oh, I'm sorry to hear that. It's okay! Tell me about your favorite movie, then. Why is it your favorite?
Where would you go if you could visit any place in the world?	Probably the UK	Wonderful! If could visit only one place in Germany, it would be…
What is the one thing you would want to do for fun in your dream destination?	I guess, um, explore like Tokyo and the mountains of Japan, and shrines	Yeah, that's cool. Actually, I forgot to mention that one of my friends is going to…
In Beetlejuice, it's so funny when everyone starts singing the "Jump in the line song!" Do you know which song I'm talking about?	No	Yeah, come on, it goes something like'shake shake shake senora shake your body line…'
I didn't know [Jared Leto] was an actor until [Alexander the Great]. Have you seen that movie? it's like 12 years old	I probably watched it a while back, but I don't remember what that was about	I saw it a long time ago, so I don't remember much of it myself, but as I was saying I first knew him…

Empirical Evaluation. We used the Paola Chat application to evaluate users' perceptions of naturalness of the agent's breathing. The study was a within-subjects design in which one of the conversations had the ECA with the breathing animations and the other conversation had the ECA that did not use the breathing animations. The design was balanced for order of breathing/non-breathing and for order of conversation topic. After each conversation, participants were asked to complete a seven-point Likert-scale survey of naturalness, rapport, and social presence.

A total of 62 participants interacted with Paola. The population consisted of college students mostly aged 18–25 (about 85%; the remaining 15% were under age 30). The population consisted of 73% males. Further, 68% of the participants were native speakers of English. Of the participants, 21 identified as first-year college students, 11 as second-year college students, 13 as third-year college students, and 7 as fourth-year college students. The remaining 10 participants were in their fifth-year of study or above.

Before the interaction, each participant was asked to complete a demographic survey. Each also signed a permission to be video-recorded during the interaction. Participants were seated in front of a wall where Paola was projected (see Fig. 1). The two conversations each lasted about five to seven minutes, with the exact length of each interaction depending on the user's responses to Paola's questions.

After the first conversation, users were asked to complete a survey on the interaction. The interaction continued with a conversation on the other topic. The session would conclude with a final survey. Table 2 displays the 18-question survey participants completed; responses were entered on a 7-point Likert scale of users' perception of naturalness, rapport and of social presence, as used, for example, in [4, 10].

Table 2. Pre-interaction and post-interaction survey questions.

Q1	I feel that the agent understood me
Q2	The agent seemed disengaged
Q3	The agent was excited
Q4	The agent's movements were unnatural
Q5	The agent was friendly
Q6	The agent was not paying attention to me
Q7	The agent and I worked towards a common goal
Q8	The agent and I did not seem to connect
Q9	I sensed a physical connected with the agent
Q10	The agent's gestures were not lively
Q11	I feel the agent trusts me
Q12	I didn't understand the agent
Q13	I perceive that I am in the presence of another person in the room with me
Q14	I feel that the person is watching me and is aware of my presence
Q15	The thought that the person is not a real person crosses my mind often
Q16	The person appears to be sentient, conscious, and alive to me
Q17	I perceive the person as being only a computerized image, not as a real person
Q18	Overall, the agent's behavior seemed natural

3 Results

The study's results suggested that breathing did not affect the agent's perceived naturalness. Table 3 displays the average scores for rapport, naturalness, and social presence. Although the average rapport scores across the experimental and control conditions were normally distributed (Anderson-Darling test, p = 033 and p = 0.26, respectively), the absolute difference in average scores was small (4.08 – 3.78 = 0.30, on a scale from 1 to 7), and a t-test was not significant (p = 0.76). Similarly the t-tests for naturalness and social presence were also not significant (both p = 0.69).

Because the design of the experiment allowed for the participants to watch Paola as she spoke or listened, the design of Paola's utterances required an emphasis on the eliciting questions. It was important to elicit longer, more thoughtful responses than

Table 3. Naturalness, rapport, and social presence on both experimental and control conditions.

	Experimental condition		Control condition	
	Mean	StDev	Mean	StDev
Rapport	4.08	1.67	3.78	1.65
Naturalness	4.13	1.59	4.24	1.49
Social presence	3.14	1.67	3.23	1.76

simple affirmations or negations. Table 4 shows questions asked by Paola that elicited longer responses from the human participants.

Table 4. Responses to questions.

Paola's question	*User's response*
Can you think of any other artists that go back and forth with movies and music?	Uh, I think someone like Justin Bieber used to do that kind of stuff. Or, I can't think of any one too much. I'm not too good on artists and that kind of stuff
	Oh, Lady Gaga. She was in the American Horror Story, I thought she did really, good there
What sorts of new movie experiences do you think will come next?	Uh, VR or they sort of already do this, but like they move the seats around and so you can feel what's going on in the movie. You can feel it in your chair
	Um, I don't know
Tell me about the vacation [you last went on]	Um, so I've gone to Washington and that was a lot of fun, that was for an interview. I've also been to Vegas, that was a lot of fun. But I wasn't 21 so I couldn't gamble, but that was a lot of fun. I also went to Florida
	Oh, it was just winter break
What is the one thing you would want to do for fun in your dream destination?	Anything, as long as it's relaxing or so I could be at the beach. Um, I don't know, just relaxed, having fun basically
	I guess, um, explore like Tokyo and the mountains of Japan, and shrines
Have you seen [Alexander the Great]? It's like 12 years old	I probably watched it a while back, but I don't remember what that was about
	No
Can you think of your favorite scene of any movie?	I don't know, maybe the fight of Star Wars 3

In responding to questions about specifics, some users chose to response briefly, with short responses such as "Oh, it was just winter break," while others gave utterances over ten words, as shown in Table 4. One of Paola's questions, about things to do in dream destinations, generated longer utterances, but these utterances tended to be more general in tone, with fewer specifics. For example, users gave answers such as "relaxing" at or "exploring" their dream destinations.

As expected, some users responded to questions requiring specific knowledge (e.g., "*Can you think of any other artists that go back and forth with movies and music?*") with expanded statements, while other questions generated one-word responses. Responses to general questions (e.g., "*What is the one thing you would want to do for fun in your dream destination?*") generated more thoughtful responses from the users, and therefore utterances with a higher word count.

Inviting users to speak about their experiences or preferences produced longer utterances, too. In both topics, users responded to questions about their favorite movie scenes or dream vacations by responding with utterances longer than six words. This was also reflected in user responses to clarifying statements by Paola, for instance, when she asked "*Tell me about the vacation*" or "*What was that movie about?*"

4 Conclusion

When we first saw SimSensei, the PTSD-diagnosis agent, we noticed that it appeared that she was not breathing. This led to us to develop the Paola Chat application, which we then used for a perception study of naturalness of breathing in the agent. Our results suggested that breathing did not actually affect perceived naturalness, rapport, or presence.

Despite our expectation that increased naturalness from breathing would lead users to report greater rapport in the breathing condition than in the not-breathing condition, the study's results suggest that animation of breathing appears to neither increase nor decrease these perceptions. Of course, while breathing by an ECA does not increase naturalness, neither does it detract from naturalness. This suggests that ECAs using a breathing model similar to that of the Paola Chat agent can be at least as natural as a non-breathing agent. When combined with other features implemented, such as generating responses not only relevant to the topic but also relevant to the utterance to which Paola responds (see Table 1), the perceived social presence of ECAs could be increased.

So why did we perceive that SimSensei was not breathing, when the participants in our study did not notice the breathing in the Paola Chat agent? One of the differences between the two agents was the amount of dialog produced by the agent. SimSensei was mainly listening to the person with whom it interacted (except for some occasional questions, nods, and hand movements), while Paola was more conversational and contributed substantively to the conversation.

A second factor may be that the Paola Chat agent's breathing was displayed only visually and not auditorily. If someone is about to speak, you can sometimes hear the inhalation.

A limitation of this study is that the Paolo Chat application was fully automated rather than Wizard-of-Oz. That is, the application generated utterances as output to users and accepted responses as input independent of a human acting behind the scenes. The dialog models were developed beforehand, where responses could be either short responses recognized by key phrases (converging to a previously designed point of the conversation) or wildcard responses (which followed the flow of the conversation more generically while being able to handle longer utterances regardless of keywords). This technology, though, constrains the agent's conversational responsiveness.

Though Paola displayed animated gestures while talking, nodding, and now breathing, the application did not include more than nominal models of gaze and head movement. Adding these kinds of animations to an otherwise static embodied conversational agent might provide an even more humanlike appearance. These features could provide an improvement in perceived naturalness because breathing affects not only the torso and neck but shoulder movement and even timing of dialog generation.

References

1. Cassell, J. (ed.): Embodied Conversational Agents. The MIT Press, Cambridge (2000)
2. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. **51**(7), 630–639 (2009)
3. Huang, L., Morency, L.-P., Gratch, J.: Virtual rapport 2.0. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS (LNAI), vol. 6895, pp. 68–79. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23974-8_8
4. Gris, I., Novick, D., Camacho, A., Rivera, D.A., Gutierrez, M., Rayon, A.: Recorded speech, virtual environments, and the effectiveness of embodied conversational agents. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS (LNAI), vol. 8637, pp. 182–185. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09767-1_22
5. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS (LNAI), vol. 6356, pp. 222–235. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15892-6_24
6. Novick, D., Gris, I.: Building rapport between human and ECA: a pilot study. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 472–480. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07230-2_45
7. Rizzo, A.A., Scherer, S., DeVault, D., Gratch, J., Artstein, R., Hartholt, A., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Stratou, G.: Detection and computational analysis of psychological signals using a virtual human interviewing agent. In: Proceedings of the International Conference on Disability, Virtual Reality and Associated Technologies (2014)
8. Włodarczak, M., Heldner, M.: Respiratory turn-taking cues. In: Proceedings of Interspeech 2016, pp. 1275–1279 (2016)
9. Novick, D., Gris Sepulveda, I., Rivera, D.A., Camacho, A., Rayon, A., Gutierrez, M.: The UTEP AGENT system. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 383–384 (2015)
10. Novick, D., Gris, I., Camacho, A., Rayon, A., Gonzalez, T.: Bigger (Gesture) isn't always better. In: Kurosu, M. (ed.) HCI 2017. LNCS, vol. 10271, pp. 609–619. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58071-5_46