An affective virtual agent providing embodied feedback in the paired associate task: system design and evaluation

Christian Becker-Asano[1], Philip Stahl[1], Marco Ragni[2], Matthieu Courgeon[3], Jean-Claude Martin[3], and Bernhard Nebel[1]

[1] University of Freiburg, Department of Computer Science, Georges-Köhler-Allee 52, 79110 Freiburg, Germany {basano, stahl, nebel}@informatik.uni-freiburg.de
[2] University of Freiburg, Center for Cognitive Science, Friedrichstr. 50, 79098 Freiburg, Germany ragni@cognition.uni-freiburg.de
[3] LIMSI-CNRS, Rue von Neumann, Batiment 508, 91403 Orsay Cedex, France {martin, courgeon}@limsi.fr

Abstract. An affective, virtual agent is presented that acts as a teacher in the classical paired associate task. It is explained, why and how the virtual agent framework MARC was combined with the cognitive architecture ACT-R, the affect simulation architecture WASABI, and the voice-synthesis module OpenMARY. The agent's affective feedback capabilities are evaluated through an empirical study, in which participants had to solve association tasks. We expected that (1) the presentation of the task by a (neutral) virtual agent would change a learner's performance and that (2) the additional simulation and expression of emotions would impact a learner's performance as well. Finally, we discuss reasons for the lack of statistically significant differences as well as planned future application scenarios of our affective agent framework.

1 Introduction

In the domain of pedagogical agents [1] it has long been claimed beneficial to equip virtual agents with the capability to both convey as well as elicit emotions. The visual quality of interactive 3D computer graphics has increased dramatically since then and also the state-of-the-art in emotion simulation has significantly advanced. Although some evidence has been gathered for particular agents in particular scenarios [2–5], answering the general question of whether or not an agent should show emotions to improve a human's performance in human-computer interaction remains an open challenge.

We set out to test the influence of affective behavior shown by a virtual agent in face-to-face interaction with a human. We used the paired associate task (introduced by [6]) to see, if the participants' performance changes, when the task is presented by an affective as compared to an unemotional agent[4]. In addition, our results are compared to those of the original study, in which the 20

[4] See also: http://www.youtube.com/watch?v=3BYTNxMs028

associations between a single monosyllabic word and a digit between zero and nine were presented as text only.

In doing so, a novel combination of several independent software components was devised as to create high-quality, convincing agent behavior. Accordingly, after an overview of related work in the next section, the system is described in Section 3. Section 4 details an empirical study together with a presentation of its results. The latter are being discussed in Section 5, in which possible directions for future research are presented as well.

2 Related work

Virtual agents are designed to capture the richness and dynamics of human behavior [7]. Different frameworks for virtual agents exist, e.g., the Virtual Human Toolkit [8], Greta [9], and MARC [10]. They differ in complexity, graphical output and application domains.

Concerning the applications of virtual humans, a study on emotional contagion between virtual and real humans [11] suggests that, although virtual agents can eliciting this effect, it is dampened when humans have to make strategic decisions at the same time. Furthermore, in a conversational setting the smiling behavior of the virtual agent MAX has been shown to be mimicked by humans [12], but its likeability was not rated higher when it smiled more often.

Taking these limitations into account, we decided to integrate our system into a task that affords only very limited interaction capabilities of our virtual human, namely, the "paired associate task" [6, 13]. It is presented by Anderson as a common task to test the capabilities of human working memory.

We extend this previous work by adding a virtual agent's emotional feedback to it. The consequences of a virtual agent's affective feedback on students have previously been investigated in a learning environment [14]. However, they focused on "empathetic feedback" that was realized by "short, text-based responses" and not, as in the system described here, in terms of facial and vocal expressions. One study investigated how positive, neutral, and negative feedback responses from AutoTutor influenced learners' affect and physiology [15]. It was found that AutoTutor's feedback correlated with the learner's affect: after positive feedback from AutoTutor, learners mostly experienced delight, while surprise was experienced after negative feedback.

3 System

Our system realizes a virtual agent with the ability to show different emotions, produce verbal output, recognize speech input and the ability to predict human input in the paired associate domain.

To this end, the following five software components were combined (cp. Fig. 1):

- The MARC framework [10] realizes the visual output by providing an affective agent, which can be controlled in real-time.

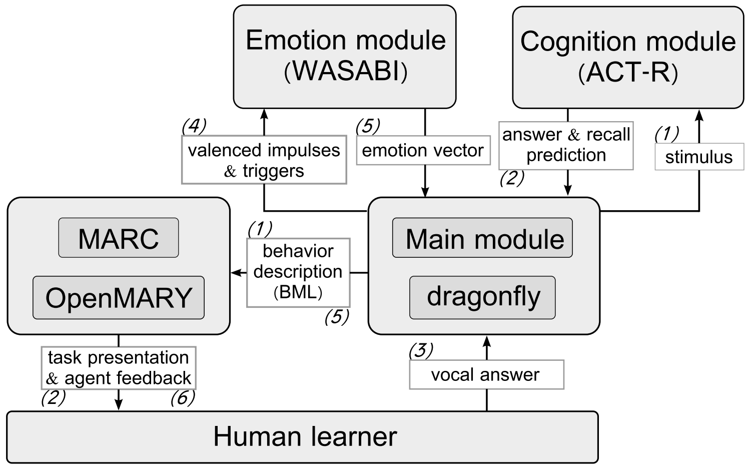

Fig. 1: System overview showing the interactions between the different modules

- The open source TextToSpeech software OpenMARY [16] generates verbal output with different affective connotations.
- The emotion module is realized by the open source, affect simulation architecture WASABI [17].
- The cognition module is a reimplementation of the memory activation function (probability of retrieval; [6]) that evaluates the likelihood to recall a specific memory chunk (and to give the correct answer). This gives an estimate of the difficulty of a retrieval.
- The main module uses Microsoft Windows 7 speech recognition via dragonfly [18] to receive user input.

In the following, the single modules and their interaction are described.

3.1 MARC and OpenMARY

MARC is a virtual agent framework for creating agents that allow real-time affective behavior. The agents' dynamic facial expressions [19] can be driven by emotions such as those provided by the WASABI architecture. The graphical output is rendered online, so that all predefined animations can be called in real-time. The agents within MARC are controlled via a dialect of the Behavior Markup Language (BML, [20]; see also *(1)* and *(5)* in Fig. 1). In combination with OpenMARY optimized lip synchronization of the rendered audio is achieved.

For the emotion mapping and the specification of animation parameters the system designer is provided with a graphical user interface. Animations can be assigned to each of the WASABI emotions, namely *Happy, Concentrated, Bored, Annoyed, Angry, Surprise, Fear, Hope, Relief* and *Fears-Confirmed*, and all their

Fig. 2: The virtual agent expressing *anger*, *neutral*, and *joy* (left to right)

MARC specific parameters like *intensity* and *interpolation* can be specified. For the empirical study reported here we limited MARC's expressions to *neutral*, *anger*, and *joy* (cp. Fig. 2), of which the latter two had been evaluated to be easily recognizable [19]. These particular emotions are also most relevant to the task in terms of influencing a student's motivation. WASABI's emotion *happy* is mapped to MARC's emotion *joy*. Furthermore, although WASABI distinguishes *annoyed* and *angry* (see Table 1), these emotions are both expressed by *anger* in the MARC framework.

OpenMARY is a open source text to speech system characterized by a modular design, an XML-based system-internal data representation and an easy to use interface which can be accessed via network protocols [16]. An important factor for achieving a realistic affective agent is the use of emotional speech synthesis. Thus, OpenMARY's German voice [21] was used to realize the affective states *anger*, *neutral*, and *happy* in the agent's vocal expressions.

3.2 WASABI and ACT-R

The agent's emotions are simulated dynamically by the WASABI affect simulation architecture [17]. Two types of input signals (cp. *(4)* in Fig. 1) are sufficient for WASABI to simulate the time course of primary and secondary emotions (see also [17, pp.148ff]):

1. *Valenced impulses* ranging from -100 to 100 are necessary to drive the internal emotion dynamics.
2. *Emotion triggers* are needed to maximize the intensity of either *angry*, *annoyed*, or *happy*. Any such maximized intensity drops off linearly to zero within ten seconds, before it is automatically reset to its predefined base intensity of 0.75.

These inputs are realized in terms of messages sent via UDP by the *main module*. WASABI, in turn, sends UDP-based messages back to the *main module* once per second containing the actual set of emotion/intensity pairs.

ACT-R functions about retrieval probability and latency are used to give the agent an understanding about the difficulty of recalling a certain item at a

Table 1: Derivation of the triggered emotion and the valenced impulses send to the emotion module depending on the discrepancy between expected and received answers. In the last column the facial expressions of the MARC agent are given associated with the WASABI emotion presented in the third column.

Expectation	human answer	emotion triggered	impulse	associated facial expr.
negative	none	none	0	neutral
negative	incorrect	annoyed	-30	anger
negative	correct	happy	80	joy
none	none	annoyed	-20	anger
none	incorrect	angry	-50	anger
none	correct	happy	50	joy
positive	none	annoyed	-50	anger
positive	incorrect	angry	-80	anger
positive	correct	happy	30	joy

certain time in the experiment. These estimations enable the agent to predict how likely the participant will give the correct answer.

When the *main module* receives the learner's answer, the *valenced impulses* are derived from this answer and the agent's expectation; cp. Table 1. If the learner's answer is correct and this was highly expected, the resulting impulse is only slightly positive and the emotion *happy* is triggered. An unexpected, correct answer, however, would also trigger *happy*, but the positive impulse would be very strong. Incorrect answers are treated in a similar fashion. The system designer can specify the mapping parameters from recall probabilities provided by ACT-R to expectation values to impulse intensities in the overall framework. The probability values are mapped to the three types of expectations as follows:

– Probabilities less than 60%: negative expectation, i.e. the answer is expected to be incorrect
– Probabilities between 60% and 90%: no expectation, i.e. the agent is unsure about which answer to expect
– Probabilities greater than 90%: positive expectation, i.e. the answer is expected to be correct

To compute the recall probabilities the base-level activation function and the probability function are reimplemented as presented by Anderson in [13, pp.74/124]. The threshold τ and the noise variable s are set to -2.0 and 0.5, respectively. These values correspond to those proposed in the ACT-R tutorial.

3.3 Interaction of the modules

The *cognition module* receives a task description from the *main module* and predicts the human learner's answer. After the learner's answer has been recognized by *dragonfly*, the *main module* triggers emotions and sends *valenced impulses* to the *emotion module*. In parallel, it listens continuously for WASABI messages containing an *emotion vector* to update the agent's affective state. The emotion with the highest intensity is encoded into a BML message to update the MARC agent's facial display accordingly. Furthermore, MARC passes on the emotion information to OpenMARY such that its synthesis is changed accordingly.

The result of the implementation is an affective agent that acts in the domain of paired associate learning as an affective tutor. Instead of providing feedback about possible error sources, it shows affective behavior according to the participant's performance hopefully further motivating him to improve performance.

4 Empirical study

We tested this system on the paired-associate task [6]. In the original study 20 association pairs are displayed for eight rounds on the screen in random order. First, the word is displayed for five seconds, then the number. From round two to eight the participants have to recall the number within five seconds after the associated word has been presented by pressing the correct key on the keyboard.

4.1 Hypothesis and Research Question

We set out to investigate two questions regarding the paired-associate task:

1. Does multimodal task presentation performed by a virtual agent change a human learner's performance?
2. How does an agent's emotional feedback impact the learner's performance, if at all?

Previous research has shown that the effects of an agent's empathetic feedback depend on whether or not a strategic task has to be solved by the human at the same time [11]. As the paired associate task is a learning and not a strategic task, we expected a significant change of the learner's performance, when this task is presented by an agent. We were unsure, however, if the integration of emotional feedback into this non-strategic task further impacts performance.

Accordingly, we established a *neutral condition*, in which the participants are expected to achieve different correctness rates than the ones reported in the original (no-agent) study. For the emotional feedback we compared a simple, rule-based approach, i.e. the *reflexive condition*, with the dynamic emotion approach realized by the integration of WASABI, namely the *WASABI condition*.

4.2 Design and Procedure

The three conditions differ in the following way:

1. *Neutral condition*: the agent shows no emotional expressions at all. This serves as a control condition testing the impact of the task presentation by an embodied agent.

2. *WASABI condition*: the emotion module is enabled. As a result the agent can change its mimic and connotation of speech according to the changing affective course in WASABI as described in Section 3.2 taking ACT-R-driven expectations into account.
3. *Reflexive condition*: the agent shows a strictly rule-based behavior by always showing an *angry* emotion in response to an incorrect answer and a *joy* emotion in response to a correct answer.

The general procedure of the paired-associate task remained the same. However, in addition to a visual presentation on the screen, the MARC agent says the word and the number at the time of their resp. presentation. In addition, our participants were requested to respond verbally from round two instead of using the keyboard. The agent then reacted to the correctness of the answer by nodding or shaking its head. The experimental setup is presented in Fig. 3, left, and the agent as it is displayed in Fig. 3, right.

Fig. 3: The experimental setup (left); the agent as it is displayed on the screen with subtitles in addition to its verbal output

4.3 Participants

Data of 60 participants were collected, of which two had to be excluded, because they did not achieve a correctness rate of 60 percent or higher in any of the seven runs (similarly to the original study [6]). The data of two more participants were removed, because they had misunderstood the instructions.

The remaining 56 participants were on average 23.3 years old with 32 of them being female. They were randomly assigned to the neutral condition (22 subjects, 14f, *mean* = 21.5 years), the WASABI condition (18 subjects, 8f, *mean* = 24.3 years), and the reflexive condition (16 subjects, 10f, *mean* = 23.7 years).

Table 2: Correctness rates (in percentages) with standard deviations for the different conditions. For the original study [6] only the overall average of the standard deviations 2.6% is available

	ORIG. STUDY		NEUTRAL		WASABI		REFLEXIVE	
Run	mean	std	mean	std	mean	std	mean	std
2	52.6	(2.6)	59.6	24.4	58.1	20.8	61.3	14.1
3	66.7	(2.6)	69.8	22.1	66.9	25.9	70.7	16.3
4	79.8	(2.6)	82.1	17.0	76.9	23.2	81.7	11.2
5	88.7	(2.6)	88.0	14.4	80.6	20.9	92.0	9.2
6	92.4	(2.6)	92.7	10.1	86.7	17.4	92.3	8.8
7	95.8	(2.6)	92.7	10.7	88.6	12.5	95.7	5.1
8	95.4	(2.6)	92.7	10.1	93.1	11.7	98.3	2.9

4.4 Results and Discussion

The results of the experiment are summarized in Table 2 comparing the values of the original study [6] with our conditions. The data sets of the conditions were analyzed by computing repeated measures ANOVAs between subjects with the statistical software *R*. All pairwise comparisons between conditions reveal no significant difference (*neutral* against *WASABI condition* $F(7, 296) = 0.112$, n.s.; *neutral* against *reflexive condition* $F(7, 272) = 0.052$, n.s.; *WASABI* against *reflexive condition* $F(7, 240) = 0.571$, n.s.). The comparison of our conditions with the original study's data has to remain on an informative level in lack of the raw data of the Anderson study. The participants' overall correctness rate in the *neutral condition* was on average 0.9% better than that reported in the original study. The correctness rates achieved in the *WASABI condition* were on average 2.9% worse than that of the original study and the one of the *reflexive condition* was 2.9% better. With regard to our research questions, the presence of an agent as teacher in the paired associate task seems not to change a learner's performance significantly. Surprisingly, a learner's correctness rates seem to benefit from an agent's affective feedback, only if it is achieved by a rather simple, reflexive emotion simulation.

A number of reasons for the insignificant differences can be speculated about:

- The task was so simple that most students reached very high recall rates early on, so that the agent's display of negative emotions could not really interfere with the memory task.
- The choice of parameters within our framework might not be optimal. The total presentation time of *joy* and *anger* was very different: In the Wasabi condition *anger* was shown on average for 46.3 seconds (3.2%) during the whole experiment, while *neutral* was presented for 493.9 seconds (34.6%) and *joy* for 887.8 seconds (62.2%). In the reflexive condition learners were on average confronted with *anger* for 212.4 seconds (14.9%), with *neutral* for

217 seconds (15.2%) and with *joy* for 998.6 seconds (69.9%). Maybe showing *anger* more often would enlarge the agent's impact on human learners.
– The learners might have been so focused on their task that an agent's affective feedback was largely ignored similar to the effects reported in [11] for strategic tasks.

5 General discussion

We set out to investigate (a) the effect of a virtual agent's presence in and active presentation of a learning task and (b) the additional effects of the agent's affective feedback during task presentation. In doing so, we implemented a new experimental agent framework that combines the state-of-the-art virtual agent framework MARC with an affective component based on WASABI, a cognitive component based on ATC-R, the voice synthesis component OpenMARY, and a voice recognition component based on dragonfly.

Although the results of the empirical study remain inconclusive, it needs to be pointed out that the experimental framework can easily be modified and extended. The following tasks will be approached next:

– Extending the number of emotions as provided by WASABI to be integrated into the simulation and displayed by the agent.
– Online emotion recognition by means of physiological sensors, facial features and/or eye tracking.
– Changing the task to one that affords more direct interaction with the agent, possibly a game like chess, e.g., similar to [22].

In conclusion, we believe that this framework can serve as a flexible test environment for further psychological studies on the effects of emotions in human-computer interaction.

References

1. Johnson, W.L., Rickel, J.W., Lester, J.C.: Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education **11** (2000) 47–78
2. Conati, C.: Probabilistic assessment of user's emotions in educational games. Applied Artificial Intelligence **16** (2002) 555–575
3. Qu, L., Wang, N., Johnson, W.L.: Pedagogical agents that interact with learners. In: AAMAS-04 Workshop on Balanced Perception and Action in ECAs. (2004)
4. Prendinger, H., Becker, C., Ishizuka, M.: A study in users' physiological response to an empathic interface agent. Intl. Journal of Humanoid Robotics **3**(3) (2006) 371–391
5. Hall, L., Woods, S., Aylett, R.: Fearnot! involving children in the design of a virtual learning environment. Intl. Journal of Artificial Intelligence in Education **16**(4) (2006) 327–351

6. Anderson, J.R.: Interference: The relationship between response latency and response accuracy. Journal of Experimental Psychology: Human Learning and Memory **7**(5) (1981) 326–343
7. Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E.: Creating interactive virtual humans: Some assembly required. IEEE INTELLIGENT SYSTEMS **17** (2002) 54–63
8. Kenny, Hartholt, Gratch, Swartout, Traum, Marsella, Piepol.: Building Interactive Virtual Humans for Training Environments. In: Interservice/Industry Training, Simulation and Education Conference. (2007)
9. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: an interactive expressive eca system. In: Proc. Intl. Conf. on Autonomous Agents and Multiagent Systems. AAMAS '09, Richland, SC (2009) 1399–1400
10. Courgeon, M., Martin, J.C., Jacquemin, C.: MARC: a Multimodal Affective and Reactive Character. In: Proc. 1st Workshop on AFFective Interaction in Natural Environements. (2008)
11. Tsai, J., Bowring, E., Marsella, S.C., Wood, W., Tambe, M.: A study of emotional contagion with virtual characters. In: The 12th International Conference on Intelligent Virtual Agents (IVA), Santa Cruz, CA (September 2012)
12. Krämer, N., Kopp, S., Becker-Asano, C., Sommer, N.: Smile and the world will smile with you – the effects of a virtual agent's smile on users' evaluation and behavior. Intl. Journal of Human-Computer Studies **71**(3) (2013) 335 – 349
13. Anderson, J., Lebiere, C.: The Atomic Components of Thougt. Taylor & Francis Group (1998)
14. Robison, J., McQuiggan, S., Lester, J.: Evaluating the consequences of affective feedback in intelligent tutoring systems. In: Affective Computing and Intelligent Interaction and Workshops, 2009. (2009) 1–6
15. Aghaei Pour, P., Hussain, M., AlZoubi, O., D'Mello, S., Calvo, R.: The impact of system feedback on learners' affective and physiological states. In: Intelligent Tutoring Systems. Volume 6094 of LNCS. Springer (2010) 264–273
16. Schröder, M., Trouvain, J.: The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology **6** (2003) 365–377
17. Becker-Asano, C.: WASABI: Affect simulation for agents with believable interactivity. Volume 319. IOS Press (2008)
18. Butcher, C.: Dragonfly speech recognition source code. http://code.google.com/p/dragonfly/ (April 2013)
19. Courgeon, M., Clavel, C., Tan, N., Martin, J.C.: Front view vs. side view of facial and postural expressions of emotions in a virtual character. In Pan, Z., Cheok, A., Müller, W., eds.: Transactions on Edutainment VI. Volume 6758 of LNCS. Springer (2011) 132–143
20. Vilhjálmsson, H., Cantelmo, N., Cassell, J., E. Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thórisson, K., Welbergen, H., Werf, R.: The behavior markup language: Recent developments and challenges. In: Intelligent Virtual Agents. Volume 4722 of LNCS. Springer Berlin Heidelberg (2007) 99–111
21. Steiner, I., Schröder, M., Charfuelan, M., Klepp, A.: Symbolic vs. acoustics-based style control for expressive unit selection. In: ISCA Tutorial and Research Workshop on Speech Synthesis (SSW-7), Kyoto, Japan, ISCA, ISCA (2010)
22. Leite, I., Martinho, C., Pereira, A., Paiva, A.: iCat: an affective game buddy based on anticipatory mechanisms. In: Proc. 7th Intl. joint Conf. on Autonomous agents and multiagent systems - Volume 3. AAMAS '08, Richland, SC (2008) 1229–1232