

Using Virtual Agents to Guide Attention in Multi-task Scenarios

Philipp Kulms, Stefan Kopp

Sociable Agents Group, Center of Excellence ‘Cognitive Interaction Technology’ (CITEC)
Faculty of Technology, Bielefeld University, Germany

{pkulms, skopp}@techfak.uni-bielefeld.de

Abstract. Humans have the ability to efficiently decode human and human-like cues. We explore whether a virtual agent’s facial expressions and gaze can be used to guide attention and elicit amplified processing of task-related cues. We argue that an emphasis on information processing will support future development of assistance systems, for example by reducing task load and creating a sense of reliability for such systems. A pilot study indicates subjects’ propensity to respond to the agent’s cues, most importantly gaze, but to not yet rely on them completely, possibly leading to a decreased performance.

Keywords: Virtual embodied assistant; attention; task-switching; social information; cognitive processes

1 Introduction and Concept

Human-machine interaction can be very complex, especially when it involves executive functions (attention allocation, priority setting, response scheduling, working memory, etc.) operating on multiple tasks. For example, in car-driving, operating industrial facilities, or air traffic control, humans often have to perform multiple tasks with time-varying demands. While Cognitive Scientists have investigated mechanisms of switching between multiple tasks and, e.g., how task interferences result in costs and errors [1], fields like human factors, software ergonomics, and human-computer interaction strive to prevent multi-task situations or, since they are often unavoidable, to provide users with support mechanisms [2]. In this paper we study how a virtual agent can provide support in such multi-task scenarios by guiding attention with social signals, and what effects the presence of such an agent has on the user and the respective tasks.

Interface agents are far from being a novel idea (see e.g. [3]). However, although positive social effects of virtual agents on processes such as learning have been confirmed [4], only a little research has investigated the detailed cognitive effects of an agent’s presence on users that have to fulfill given tasks. We focus on situations where users have to carry out multiple tasks simultaneously (see Fig. 1, A). In most cases, these tasks interfere in some way and automatic performance is hard to estab-

lish. This means users have to employ a task-switching strategy that requires to manage attention accordingly: they would focus on one task temporarily, but have to monitor the other to decide whether a switching is necessary (Fig. 1, B). This split attention as well as frequent task switching is likely to hamper performance in both tasks. We explore whether a virtual agent can assist in this task-switching by guiding attention to those tasks that need to be attended to (Fig. 1, C).

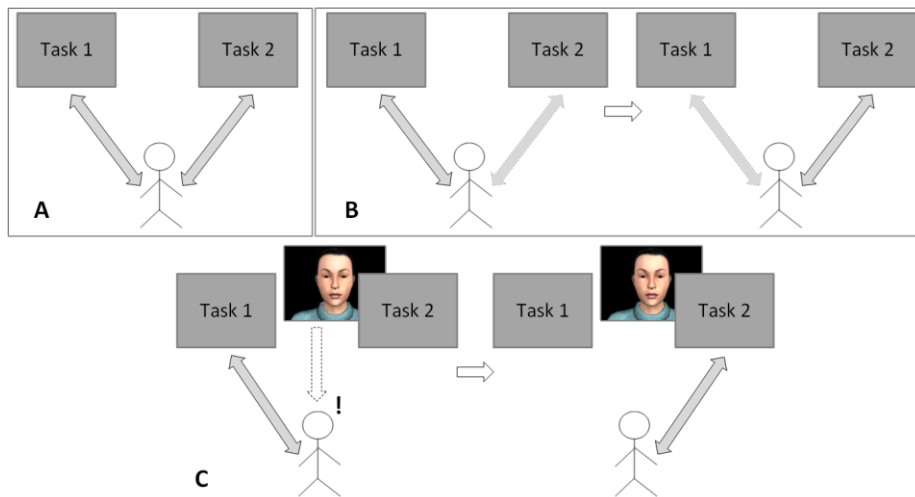


Fig. 1. Schematic rationale of our virtual assistant. In most situations, performing multiple tasks truly simultaneously is virtually impossible (A). Thus the user has to switch between the tasks, based on interlaced observation of both tasks which may hamper the performed task (B). We envision an agent to socially signal the need to switch tasks in resource-sparing ways (C).

Our approach is based on two assumptions: First, virtual human-like agents are social actors, i.e. they are able to issue social signals perceived as such, and a variety of research suggest that socially and emotionally salient cues can modulate selective attention, working memory, and motor responses (for an overview, see [5]). In particular, social stimuli appeal to human low-level perception and automatic, efficient bottom-up mechanisms of perception that amplify identification of environmental cues and release resources for executive processes, including planning and decision making [6]. For example, our brain is optimized to categorize facial expressions and efficiently de-correlates the basic expressions happiness, surprise, fear, anger, disgust, and sadness, such that overlapping signals are reduced [7]. The second assumption is that attention is controllable top-down, can be geared toward execution of specific behaviors, and is easier to divide in situations of low task load [8]. Such top-down attention mechanisms can facilitate visual attention for positive or arousing emotional stimuli [5]. Similar facilitation effects can be observed for the social information a face conveys, and executive functions were recently shown to be burdened less by social cues, pointing to a working memory unit for social information [9].

Clearly, gaze is another highly task-relevant cue. Friesen and Kingstone [10] have shown that people respond significantly faster to a target when gaze direction by a schematic face predicted the target location. This finding was confirmed for facial attention cues when additional head movements were presented to the observer which suggests that orienting toward another person's social attention is reflexive and stimulus-driven [11]. Within human-computer interaction, gaze-based attention regulation warrants interactive goal-directed communication [12, 13]. Naturally, from a conversational point of view, researchers are very much interested in cues that shape speaker/listener behaviors, one of which, of course, is gaze [e.g. 14, 15]. However, since gaze is closely tied to both sensing and communicating attention, it is in fact qualified for regulating information processing in a variety of scenarios [16].

Against this background, we want to explore whether a virtual agent that is additionally present in a multi-task scenario can assist users with managing attention and task-switching, such that less cognitive resources are expended on the switching and more are available for the tasks. To this end, the agent should facilitate task performance and display easy-to-perceive social behaviors that help to guide attention towards specific (sub-)tasks. Specifically, we are interested in the orienting and alerting dimension of attention [17] and whether an agent's social cues can provide better support for guiding attention during multi-tasking compared to non-social cues. We utilize emotional facial expressions and gaze as examples of such social cues that trigger robust, automatic bottom-up response mechanisms, bias attention, amplify processing, and can be facilitated top-down. In a related study on a virtual agent's ability to cue observer attention, subjects responded more quickly to fully animated gaze cues than to static or non-fully animated images [18]. For the virtual agent Greta, a facial expression taxonomy was created which lists "I warn" as one possible functional value, involving tense eyelids, a small frown, and eye-contact [19]. In the following we present an initial pilot study exploring the effects of a virtual agent with similar social cues vs. simple graphical symbols to guide the user's attention in a multi-task scenario. Such social cues will usually be very different from the task-related information such that an additional, non-interfering channel of easy-to-process information is opened. The results are discussed as first evidence and basis for future studies.

Obviously, a number of open questions arises. For one thing, pursuing a specific task-related goal can engage different cognitive processes that might interfere with the additional "task" of observing an agent, albeit the task-switching may be lightened. As a potential issue, distraction may arise if the agent's behavior cannot be decoded efficiently during task performance but instead leads to confusion or frustration.

2 Pilot Study

As a pilot study ($N = 10$), we conducted a dual-task experiment with a cognitively demanding primary task and a secondary task that does not require continuous but occasional attention. In the agent condition (see Fig. 2), a virtual agent was displayed

that exhibits cues which relate to the status of the secondary task. In the control condition, basic graphical symbols fulfilled this function. All subjects were right-handed. They were randomly assigned to one of the conditions.

Procedure. Subjects were asked to perform both tasks at the same time and to consider any assisting cues for their actions. The primary task was a semantic task including word categorization (noun, verb, or adjective) by clicking the appropriate response button. The number of correct answers was displayed at the bottom of the window. The secondary task was to monitor a glass getting filled with water and try to prevent it from overflowing by pressing a 'Drain' button. The water level was not visible unless a 'Status' button was pressed and held. Since subjects were occupied with the primary task as well, they were required to find a strategy which enables them, instead of constantly switching between both tasks, to monitor the secondary task efficiently while performing the primary task. When the glass overflowed due to delayed draining, it gets replaced by an empty one, a warning light went on, and a score displayed how often this occurred. In order to ensure that the secondary task's status cannot be predicted, the rate of filling varied pseudo-randomly. Furthermore, to prevent subjects from blindly draining the water as often as possible, they were told that draining comes with a cost which, at some point, increases the difficulty. Subjects were told that the goal is to maximize the score of the primary task and to keep the negative score as well as the costs of the secondary task as low as possible.

Prior to the main run, subjects went through a short training phase. Shortly before the end of the experiment, subjects were asked to comment what they perceived on the screen and the way they perform their tasks, without letting their focus drift away from their actual performance. A brief acoustic signal was used as start signal. Comments were audio-taped for later analysis. The input and output channels were unimodal; users relied on visual perception and responded via mouse click. Each session lasted approximately 10 minutes.

Manipulations. The agent exhibited fully animated social cues. The status of the secondary task was linearly mapped onto the intensity of a surprised and thus alerting facial expression (eyebrows and eyelids raised, mouth open; see Fig. 2). That is, the more the glass is filled with water the more intense the expression. Additionally, a few moments before the glass is full, the agent gazes at its direction. Virtual agent behavior was generated with *AsapRealizer* [20]. In the control condition, a colored elliptic shape of the size of the agent's face replaced the facial expression. The coloring ranged from green (empty glass) to dark orange (full glass). Instead of gaze, an animated arrow of the size of the agent's head was used. Since these items usually are well-known, we expected them to be clear attentional cues.

Dependent Variables. In the dual-task study, primary task performance was assessed in terms of the number of correct and erroneous word categorizations. Secondary task performance was assessed using the number of overflowing glasses and how often subjects checked the status instead of relying on the assistant's cues.

Evaluation of Facial Expressions. An online study ($N = 95$) was conducted to evaluate the facial expressions of the agent. Above all, the study had the purpose of confirming that the agent's alerting face is reliably recognized. Subjects were presented two still pictures consecutively, showing the agent's neutral and alerting face, respec-

tively. The pictures were rated on a 5-point Likert scale. A paired-samples t-test indicates that the alerting face ($M = 4.28$, $SD = .93$) was perceived as significantly more alerting than the neutral face ($M = 2.08$, $SD = 1.24$), $t(95) = 15.68$, $p < .001$. Moreover, the alerting face was attributed significantly stronger with conveying information about the environment, that is, giving a hint about something (alerting: $M = 2.94$, $SD = 1.32$, neutral: $M = 2.06$, $SD = 1.12$), $t(95) = 5.93$, $p < .001$, and carrying an important information (alerting: $M = 2.79$, $SD = 1.35$, neutral: $M = 1.92$, $SD = 1.04$), $t(95) = 5.93$, $p < .001$. Interestingly, the alerting face ($M = 1.86$, $SD = 1.14$) was perceived as carrying less worry/concern than the neutral face ($M = 3.17$, $SD = 1.25$), $t(95) = 8.16$, $p < .001$. No difference occurred for the attribution of a sense of danger to the faces. Note that the stimulus material only depicted the facial expressions and that rating instructions did not point to anything the agent actually or hypothetically observes.

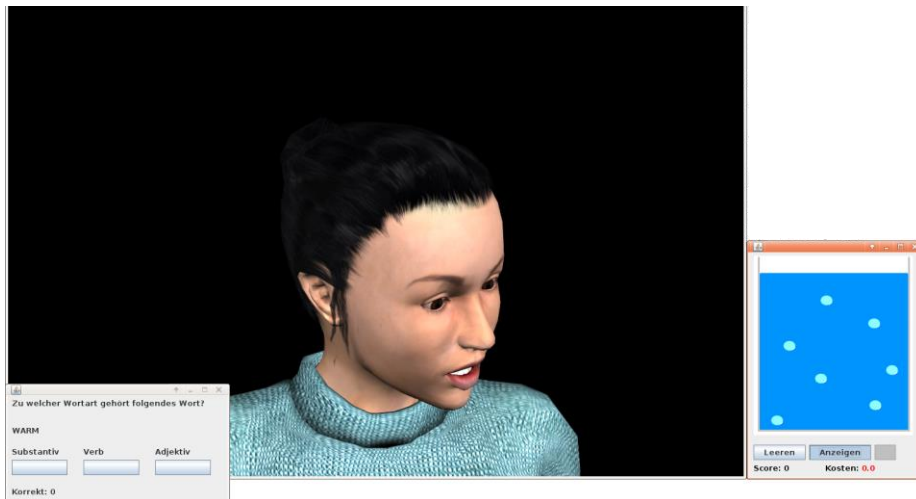


Fig. 2. Snapshot of the system in the agent condition (left: primary task; right: secondary task). The agent gazes with an alerted facial expression toward the secondary task as the glass is almost full, demanding for user reaction.

3 Results

It is important to note that the assessed data merely serve as indicator of whether subjects are able to relate agent behavior to the task and whether a tendency to rely on it can be found. Obviously, the small dataset does not qualify for statistical inference and exhibits considerable standard deviations. The primary task error rate of one subject was excluded from the dataset due to performing worse than three times the standard deviation.

Results of the primary task indicate lower performance in the presence of the agent. On average, subjects categorized 303.20 words correctly ($SD = 69.45$) which is somewhat less compared to the control condition ($M = 365.20$, $SD = 116.16$). Error

rates did not differ (agent: $M = 20.00$, $SD = 12.10$; control: $M = 20.75$, $SD = 8.23$). Performance in the secondary task too was somewhat worse under assistance of the agent as subjects let the glass overflow more often (agent: $M = 18.40$, $SD = 6.91$; control: $M = 16.80$, $SD = 3.49$). Finally, subjects checked the status more often when they were guided by the agent ($M = 53.00$, $SD = 30.54$) than by symbols ($M = 24.60$, $SD = 15.85$).

Analysis of the qualitative data suggests that subjects interpreted gaze of the agent as alerting signal and felt inclined to respond to it. The same holds for the arrow in the control condition, yet the combination of arrow and colored shape caused occasional confusion because some tried to connect the symbols' meanings with each other. All subjects reported relying on gaze of the agent entirely, ignoring or simply not perceiving its facial expressions.

4 Discussion

In both tasks the same pattern seemed to emerge: control condition subjects who were guided by symbols instead of social cues, performed somewhat better than agent condition subjects. They categorized more words correctly, did not make more categorizing errors, and managed the water level of the glass more successfully. These results do not seem in line with the fact that the agent's alerting facial expression was correctly identified in the online study and was attributed with a signaling function. Furthermore, subjects perceived the agent's gaze as an attentional cue and intuitively followed it attending the secondary task. So, why did participants in the agent conditions perform worse than in the control condition? A number of explanations seem possible.

The most likely explanation is that in the condition with human-like alerting cues, subjects allocated more attention to the hidden status of the secondary task. That is, instead of relying on the agent as a prompting device for the secondary task, subjects may have adopted more complex and time-consuming behavioral sequences consisting of checking and reacting operations. Since subjects apparently perceived gaze as an attentional cue, this means they went halfway toward our envisioned goal (Fig. 1). The downside is that, despite subjects' propensity to respond to the agent's cues and to use the agent as attention regulation device for the secondary task, they may have not relied on them completely. One possible reason could be that subjects did not yet "trust" it enough to switch tasks solely based on its behavior. It will be important to further explore this hypothesis against the background of potential issues mentioned in the introduction.

Another explanation is suggested by subjects' thoughts on their own proceeding throughout the tasks. They indicated that in a demanding situation, social cues of the agent have to exhibit a certain saliency to be recognized, despite the fact that they convey task-relevant information. It may thus be helpful to track user gaze in order to unravel the flow of information processing during the interaction and to check whether the cues themselves were attended to. Indeed, the results on interpretation of the facial expressions' emotional content point to the need for careful investigation of

facial expressions, their saliency, and unequivocal interpretation in future studies. Furthermore, on the appearance level of the agent, even its gender may significantly alter attention regulation, at least during initial phases [21].

Finally, we cannot rule out the possibility that the mere social presence of the agent may have had a negative influence on primary and secondary task performance. That is, the presence of the agent with its human-like cues may have caused distraction or irritation, thus hampering task performance. Designing more natural interactions and inducing a stronger sense of a shared goal may at least minimize the social presence issue.

In sum, our study has revealed important and interesting findings on how social cues (here, gaze and facial expression in combination) are processed in cognitively demanding settings. The next steps will be to put increased emphasis on agent cues and users' perception of said cues, so as to create a sense of reliability and trust, arising either from task interaction or pre-task priming. While the fundamentals of information processing in cognitively demanding scenarios are laid out [22], we now seek to explore the context-dependent factors that influence the processing of virtual agent cues in such settings. This may lead to new insights in how human-like cues delivered by a virtual agent are processed by humans, and which of these cues can be exploited to support users in multi-task performance.

Acknowledgements. This research was supported by the German Federal Ministry of Education and Research (BMBF) within the Leading-Edge Cluster 'it's OWL', managed by the Project Management Agency Karlsruhe (PTKA), as well as by the Deutsche Forschungsgemeinschaft (DFG) within the Center of Excellence 277 'Cognitive Interaction Technology' (CITEC). The authors are responsible for the contents of this publication.

References

1. Monsell, S.: Task switching. *Trends in Cognitive Sciences* 7(3), 134–140 (2003)
2. McFarlane, D.: Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human-Computer Interaction* 17, 63–139 (2002)
3. Lieberman, H.: Autonomous interface agents. In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pp. 67–74. ACM, New York (1997)
4. Mayer, R.E., DaPra, C.S.: An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology* 18(3), 239–252 (2012)
5. Pourtois, G., Schettino, A., Vuilleumier, P.: Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological psychology* 92(3), 492–512 (2012)
6. Yantis, S.: Goal-directed and stimulus-driven determinants of attentional control. In: Monsell, S., Driver, J. (eds.) *Attention and performance XVIII*, pp. 73–103. MIT Press, Cambridge (2000)
7. Smith, M.L., Cottrell, G.W., Gosselin, F., Schyns, P.G.: Transmitting and Decoding Facial Expressions. *Psychological Science* 16(3), 184–189 (2005)
8. Kahneman, D.: *Attention and effort*. Prentice Hall, Englewood Cliffs (1973)

9. Thornton, M.A., Conway, A.R.A.: Working memory for social information: Chunking or domain-specific buffer? *NeuroImage* 70, 233–239 (2013)
10. Friesen, C.K., Kingstone, A.: The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review* 5(3), 490–495 (1998)
11. Langton, S.R., Bruce, V.: Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, 6(5), 541–567 (1999)
12. Chopra-Khullar, S., Badler, N.I.: Where to look? Automatic attending behaviors of virtual human characters. In: Etzioni, O., Müller, J.P., Bradshaw, J.M. (eds.) *Proceedings of the Third International Conference on Autonomous Agents*, pp. 16–23. ACM, New York (1999)
13. Peters, C., Asteriadis, S., Karpouzis, K.: Investigating shared attention with a virtual agent using a gaze-based interface. *Journal on Multimodal User Interfaces* 3(1–2), 119–130 (2010)
14. Lee, J., Marsella, S., Traum, D., Gratch, J., Lance, B.: The Rickel gaze model: A window on the mind of a virtual human. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) *Intelligent Virtual Agents. LNCS*, vol. 4722, pp. 296–303. Springer, Berlin, Heidelberg (2007)
15. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., Hagita, N.: Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In: *HRI '09 Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 61–68. ACM, New York
16. Shell, J.S., Selker, T., Vertegaal, R.: Interacting with groups of computers. *Communications of the ACM* 46(3), 40–46 (2003)
17. Posner, M.I., Petersen, S.E.: The attention system of the human brain. *Annual Review of Neuroscience*, 13(1), 25–42 (1990)
18. Martinez, S., Sloan, R., Szymkowiak, A., Scott-Brown, K.: Using virtual agents to cue observer attention. In: *CONTENT 2010: The Second International Conference on Creative Content Technologies*, pp. 7–12. International Academy, Research, and Industry Association (2010)
19. Pelachaud, C., Poggi, I.: Subtleties of facial expressions in embodied agents. *The Journal of Visualization and Computer Animation* 13(5), 301–312 (2002)
20. van Welbergen, H., Reidsma, D., Kopp, S.: An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) *Intelligent Virtual Agents. LNCS*, vol. 7502, pp. 175–188. Springer, Berlin, Heidelberg (2012)
21. Mojzisch, A., Schilbach, L., Helmert, J.R., Pannasch, S., Velichkovsky, B.M., Vogeley, K.: The effects of self-involvement on attention, arousal, and facial expression during social interaction with virtual others: A psychophysiological study. *Social Neuroscience* 1(3–4), 184–195 (2006)
22. Wickens, C.D., Carswell, C.M.: Information processing. In: Salvendy, G. (ed.) *Handbook of Human Factors and Ergonomics*, pp. 111–149. John Wiley & Sons, Hoboken, NJ