Animated Pedagogical Agents Performing Affective Gestures Extracted from the GEMEP Dataset: Can People Recognize Their Emotions?

Magzhan Mukanova[1], Nicoletta Adamo[1(✉)], Christos Mousas[1], Minsoo Choi[1], Klay Hauser[1], Richard Mayer[2], and Fangzheng Zhao[2]

[1] Purdue University, West Lafayette, IN 47907, USA
nadamovi@purdue.edu
[2] University of California Santa Barbara, Santa Barbara, CA 93106, USA

Abstract. The study reported in the paper focused on applying a set of affective body gestures extracted from the Geneva Multimodal Emotion Portrayals (GEMEP) dataset to two pedagogical animated agents in an online lecture context and studying the effects of those gestures on subjects' perception of the agents' emotions. 131 participants completed an online survey where they watched different animations featuring a female and a male animated agent expressing six emotions (anger, joy, sadness, disgust, fear, and surprise) while delivering a lecture segment. After watching the animations, subjects were asked to identify the agents' emotions. Findings showed that only one expression of the angry emotion by the female agent was recognized with an acceptable level of accuracy (recognition rate >75%), while the remaining emotions showed low recognition rates ranging from 1.5% to 64%. A mapping of the results on Russel's Circumplex model of emotion showed that participants' identification of levels of emotion arousal and valence was slightly more accurate than recognition of emotion quality but still low (recognition rates <75% for 5 out of 6 emotions). Results suggest that hand and arm gestures alone are not sufficient for conveying the agent's emotion type and the levels of emotion valence and arousal.

Keywords: Affective Body Gestures · Animated Pedagogical Agents · GEMEP database · Emotion Recognition

1 Introduction

The Internet is replete with free online instructional videos featuring an instructor next to a progression of slides, or the instructor is portrayed as a talking head next to a progression of slides, or even in which the instructor provides voice over for a progression of slides. Although the script and slide content of instructional videos may be informationally appropriate, their effectiveness can be diminished because the instructor delivers them in a way that lacks emotional and social sensitivity for the learner. Preliminary research shows that the effectiveness of instructional video depends both on cognitive factors

A. L. Brooks (Ed.): ArtsIT 2023, LNICST 565, pp. 271–280, 2024.
https://doi.org/10.1007/978-3-031-55312-7_20

(such as the content and organization of the presentation) and affective and social factors, such as the degree to which the instructor displays certain emotions via voice, facial articulations and body movements, and the degree to which the learners feel rapport with the instructor (Fiorella, in press; Mayer, 2020; Mayer, Stull, & Fiorella, 2020).

Our research seeks to create an AI-based animation system that takes as input video lectures with human instructors and converts them into animated presentations delivered by affective agents. The pedagogical agents show different emotional styles and social intelligence intended to maintain a connection with the learner.

The research outlined in the paper represents progress toward achieving this objective, as it focused on how agents' emotions can be conveyed through body cues. More specifically, our study investigated whether people could recognize the quality, valence, and arousal of lecturing animated agents' emotions from a set of body gestures (no voice and facial articulations), which were extracted from the GEMEP video dataset (Bänziger & Scherer, 2010), converted to animation data, and applied to the pedagogical agents.

The paper comprises 5 sections: Section 1 is an introduction to the research; Section 2 includes an analysis of prior work on animated pedagogical agents, expression/perception of emotion in animated agents and emotion theories. Section 3 describes the study and Sect. 4 reports the study findings. Discussion and conclusion are presented in Sect. 5.

2 Review of Prior Work

2.1 Animated Pedagogical Agents

Prior research suggests that pedagogical agents (APAs) could lead to increased students' learning (Johnson & Lester, 2016; Martha & Santoso, 2019; Schroeder et al., 2013). Research also indicates that APAs have the potential to improve students' perceptions of online courses (Annetta & Holmes, 2006), while employing multimodal agents can result in more effective learning compared to agents utilizing a single channel (Cui et al., 2017).

Poggiali's study (2018) revealed that students favored "animated videos with agents for easier learning experiences, largely due to enhanced attention retention," highlighting the impact of the agent's personality in increasing engagement levels. Similarly, Mayer and DaPra (2012) observed that students exhibited improved learning outcomes when instructed by a fully embodied agent employing human-like gestures and an appealing voice, compared to an agent lacking these human-like communication elements. Furthermore, Lawson et al. (2021) established that students could identify and react to the emotional expressions conveyed by agents in a STEM-focused video lecture. Other experiments show that the agent's voice and look (Fiorella & Mayer, in press; Mayer, 2021), visual presence (Rosemberg-Kima et al., 2008), non-verbal communication (Baylor & Kim, 2009), and communication style (Wang et al., 2008) can impact learning and motivation. Pedagogical agents equipped with emotional capacities can facilitate simulated social interactions between learners and computers. A comprehensive meta-analysis, which examined results from various studies evaluating the efficacy of affective agents in educational technology-based settings, revealed a substantial and

moderate influence of incorporating emotions into APAs on students' ability to stay motivated and to retain and apply the acquired knowledge to other domains (Guo & Goh, 2015).

2.2 Expression/Perception of Emotion in Animated Agents

People use different channels, such as facial expressions, body postures, dynamic gestures, and voice to express emotions. Nonverbal cues play a significant role in communicating emotions and have important functions such as conveying a person's personality, feelings, and attitude (Meyer et al., 2021; Cheng et al., 2019). All these emotional channels are crucial in building an animated agent's personality and creating emotional bonding with the students. Research shows that body gestures can be responsible for up to 93% of the conversation (Larsson, 2014). Among body movements, the perception of emotions from hands only or just one hand was more precise than heads, torso, or arms (Blythe et al., 2023). A study by Ross & Flack (2020) revealed that images without hands impeded recognition of angry and fearful emotions.

Although some studies indicate that basic emotions can be conveyed through body movements and static poses alone, it is not easy to detect emotions solely from these cues. Certain emotions like anger, sadness, and happiness are easily recognizable through body gestures (Karg et al., 2013). However, conveying emotions such as surprise, disgust, and fear from arm movements alone poses significant challenges (Sawada et al., 2003). Atkinson et al. (2004) found that sadness and disgust were frequently misinterpreted for each other. Additionally, Ennis et al. (2013) observed that relying solely on body gestures, without facial expressions, led to inaccuracies in distinguishing emotions with a high level of arousal. Ennis concluded that while identifying happy and angry gestures was challenging, differentiating between sadness and fear was more feasible. Karg's argument emphasizes that conveying emotions with positive valence and low arousal, like contentment, was particularly difficult using gestures alone (Karg et al., 2013). Although gestures might aid in distinguishing between emotions based on valence, movement characteristics appeared to be more effective in conveying the emotion arousal level compared to static poses.

2.3 Emotion Theories/Classifications

There are two dominant emotion theories: the basic emotion theory (Ekman, 2003; Ekman & Friesen, 1969) and the dimensional theory (Russell, 2003). These theories diverge significantly, mainly concerning whether emotions are discrete entities or rather defined by two continuous dimensions—(1) valence, spanning from positive to negative, and (2) arousal, ranging from activation to deactivation. According to basic emotion theory (Ekman, 2003), humans exhibit a small number of emotions (e.g., fear, anger, joy, sadness, surprise, disgust), each displaying a consistent pattern of associated behavioral components.

In contrast, Dimensional Emotion Theory, as proposed by Russell (1980; 2003), asserts that all emotions can be positioned within a circular framework known as the circumplex of emotions, governed by two distinct dimensions: hedonic/valence (pleasure-displeasure) and arousal/engagement (rest-activated) (Russell, 1980; Posner et al., 2005;

Barrett and Russell, 2015). Within this circumplex, the horizontal axis denotes hedonic valence, while the vertical axis denotes arousal levels. Consequently, each emotion's placement within the quadrant reflects varying degrees of hedonic (valence) and arousal attributes (Posner et al., 2005). The study detailed in the paper considered both Ekman's and Russell's classifications of emotion.

3 Description of the Study

The goal of the study was to examine the extent to which the participants were able to identify six emotions displayed by two APAs performing a set of affective body gestures extracted from the GEMEP video dataset. The APAs performed the gestures while delivering a science lecture segment; the agent's voice was muted, and the face was blurred to prevent the subjects from seeing the agent's facial articulations. After watching each animated clip, the subjects were prompted to input the emotion displayed by the agent. The study used a within-subject design. The independent variables were the affective gesture type and the agent's gender; the dependent variable was the participant's emotion recognition.

3.1 Stimuli

The GEMEP database of affective expressions consists of 145 videos of French speaking professional actors portraying 12 different emotions. Each emotion is conveyed by 10 expressions, with each expression performed by a different actor. Two of the 12 emotions, e.g., surprise and disgust, are portrayed using 5 expressions only. Our study focused on six basic emotions, namely anger, fear, sadness, joy, surprise, and disgust. Only four of the most expressive recordings per emotion, two for each actor's gender, were selected from the GEMEP dataset for use in our experiment. The selection of the recordings to be used in the study was done by a group of five experts in animation and acting.

Video-based motion capture software (Deepmotion) was used to extract the actors' body gestures from the videos and convert them to animation data (Fig. 1, left). Autodesk Maya was used to clean up the motion captured data, and Unity cross-platform game engine to apply the affective gestures to the animated lecturing agents. As the videos in the GEMEP database were recorded in a professionally lit environment with a black background and clearly visible actors in the foreground, the animations of arm and hand movements were extracted with acceptable quality. Due to the fast movements of some expressions, the software could not detect the speed and all micro movements fully. To solve this issue, the videos' speed was decreased to 50% to better detect the fine and fast motions. The extracted gestures from the slowed-down videos were more accurate, however some manual editing of the animation data, especially hand and finger motions, was necessary. After the manual corrections, the affective gesture animations were applied to two of the lecturing animated agents produced by Adamo et al. (2022). The duration of each animation clip varied from 2 to 3 seconds, and to have the same length of animations for all stimuli, only 5 seconds of lecturing animation were used. As the study focused on expression of emotions through body gestures alone, the face of the animated agents was blurred, and the voice was muted. To allow for better perception

of the playing gestures, two views (front and side views with slight tilt) were included in the stimuli animations. Figure 1 (right) shows two frames extracted from two of the stimuli.

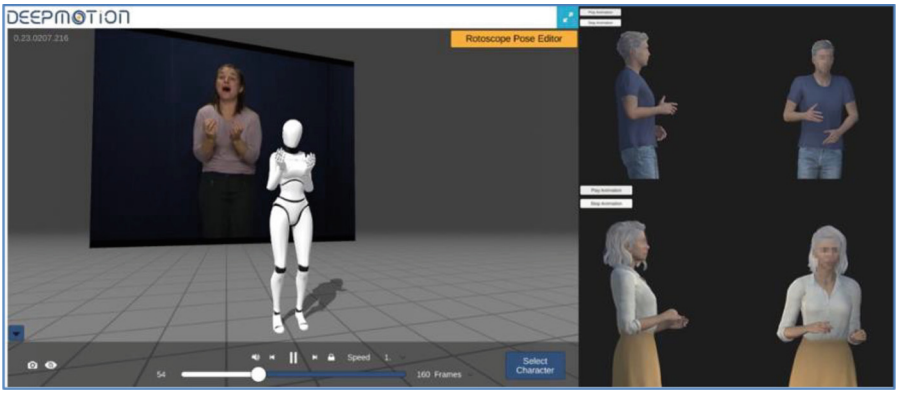

Fig. 1. Left: screenshot from video-based motion capture software showing the video of the actor in the background and the extracted gesture applied to a 3D mannequin in real- time. Right: frames extracted from two of the animation stimuli.

3.2 Subjects

205 subjects participated in the online survey, but only 131 survey responses were complete and therefore considered for further analysis. The age of the participants ranged between 16–53 years old, but most of the subjects were 18–35 years old (92%). 52% of the subjects identified themselves as female, 44% as male, 3% as non-binary and one person declined to declare their gender. Most participants indicated they were Asian, (54%) and White (37%). People with nationality from the United States of America (38%), Kazakhstan (34%), India (8%), and China (8%) made up the majority of the participants.

3.3 Procedure and Evaluation Instrument

The subjects were recruited via email. Those who were willing to take part in the study were given access an online questionnaire. The survey included 26 animation video stimuli which were presented in random order (4 animations x 6 emotions + 2 animations without affective gestures). After watching each video, the subjects were prompted to detect the emotion displayed by the agent. The subjects could select the emotion from a drop-down list with 7 items: joy, sadness, anger, disgust, fear, surprise, other. If the participants selected "other", they could input the perceived emotion. The authors used the feeling wheel to categorize the emotions entered in the text box. The last screen of the survey included a set of demographic questions.

4 Results

Figure 2 shows the emotion recognition rate for each emotion expression. To consider an emotion as accurately recognized, the authors set a recognition rate of 75% or higher based on prior studies on emotion recognition. Only one expression passed the 75% bar. It was an angry emotion expressed by the female agent, which was correctly recognized by 93.1298% of the participants; the remaining animations did not pass the set bar. The second emotion with the highest recognition rate was still anger expressed by the male agent (64.12% recognition rate). Overall, the most recognized emotions were anger (37.4046%–93.1298%), fear (the highest was 58.0153%), and sadness (58.7786%). The emotions recognized with the lowest level of accuracy were surprise and sadness (recognition rate <23% overall).

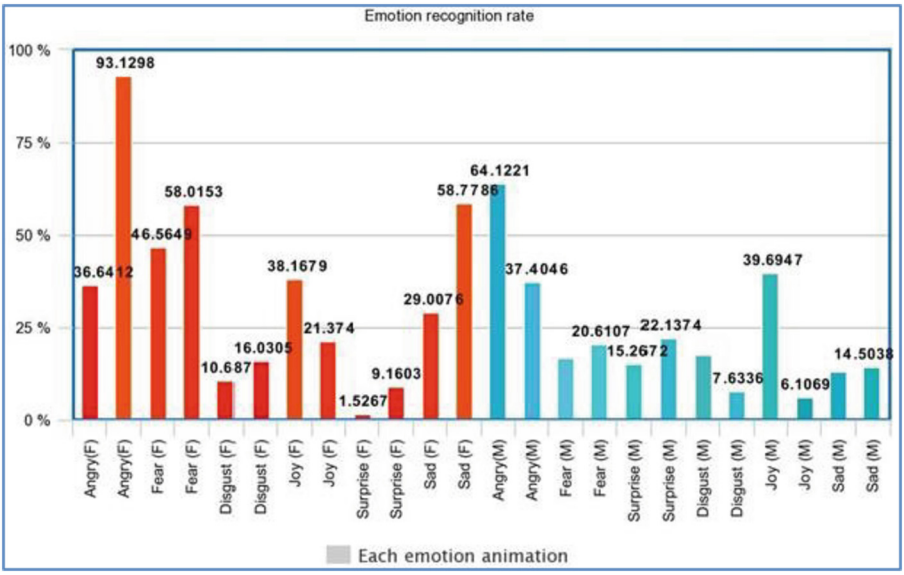

Fig. 2. Bar chart showing the emotion recognition rates. Each emotion had 4 expressions (two expressions performed by the male agent and two performed by the female agent)

Participants recognized the emotions displayed by the female agents more accurately than the emotions displayed by the male agent. However, results of pair sample t-tests comparing each emotion recognition rates for the female versus the male agent showed that the differences were not statistically significant at a significance level set to 0.05 (all p-values >0.05).

A mapping of the results on Russel's Circumplex model of emotion, showed that participants' identification of levels of emotion arousal (high versus low) and valence (negative versus positive) was more accurate than recognition of emotion quality (see Fig. 3). The identification of the levels of arousal and valence was considered correct if the emotion selected by the participant belonged in the same quadrant as the emotion

portrayed by the agent. In other words, if the agent was expressing an angry emotion, which has high arousal and negative valence (top left quadrant) and the participant, for instance, input 'stressed' as the perceived agent's emotion, the levels of valence and arousal were considered correct as the 'stressed' emotion belongs in the top left quadrant as well. For two expressions of anger more than 75% of the participants identified the correct levels of valence and arousal: Overall, fear and joy were the second highest, followed by sadness, disgust and surprise respectively.

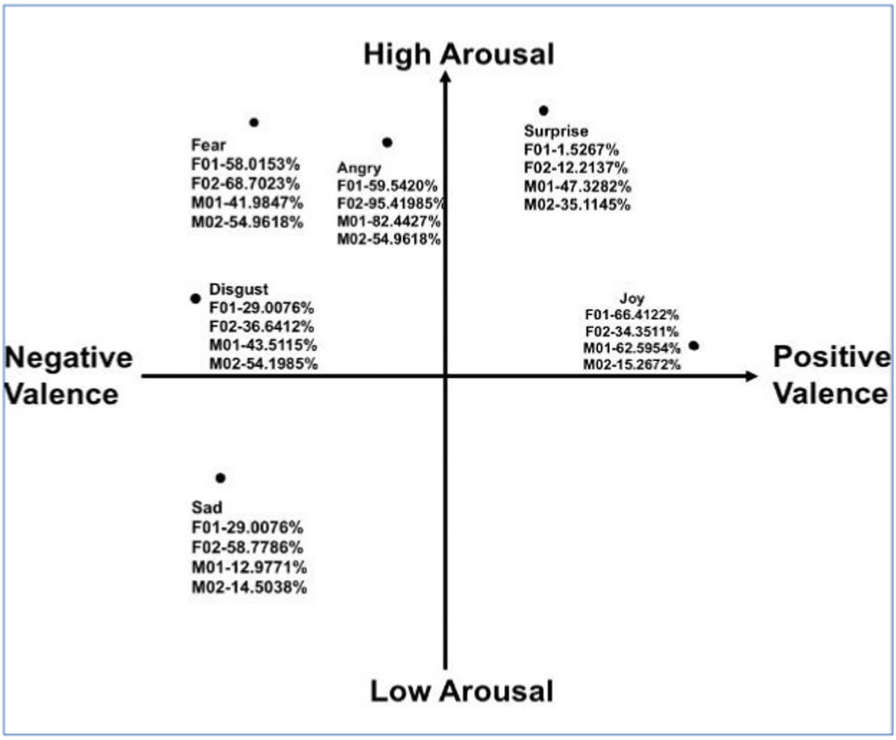

Fig. 3. Percentage of participants who recognized the correct levels of arousal and valence for each of 24 expressions of emotions.

5 Discussion and Conclusion

A significant discussion has emerged regarding whether body poses and movement effectively communicate emotions or merely indicate the intensity of those emotions (Ekman & Friesen, 1969; Lhommet & Marsella, 2015). The findings from the study suggest that, except for one emotion, hand and arm gestures alone are not sufficient for conveying the agent's emotion type or for expressing the correct levels of emotion valence and arousal. Our results align in part with prior work that showed that some emotions, such as anger, can be detected more easily from body cues (Karg et al., 2013) while

other emotions, such as surprise, disgust, and fear are the most challenging emotions to convey from arm motions alone (Sawada et al., 2003). In our experiment, anger had the highest recognition rate, whereas disgust and sadness were the most difficult emotions to detect. Further, the results of our study support prior research which indicates that people tend to perceive the emotions conveyed by female agents more accurately than those displayed by male agents.

The study had some limitations that might have affected the results. First, our experiment considered a limited set of affective gestures selected from the GEMEP dataset, hence the results may be due to the inherent features of those gestures. A different group of affective gestures could, for instance, yield higher emotion recognition rates, so it is not possible to state with certainty that body gestures alone are not sufficient for expressing animated agents' emotions.

Second, the affective hand and arm gestures in the GEMEP video database were performed by French actors while most of the subjects were white Americans and Asians. Thus, cultural differences in perception of body gestures might have affected the results. Future experiments should include a larger set of affective body gestures extracted from different datasets and a larger and more diverse pool of subjects. Third, only two animated pedagogical agents were considered in the study. It is possible that the results are dependent in part on the design features and body proportions of the agents. Future research should examine whether the look and visual style of the agents have an effect on perception of the affective gestures.

The findings from our study are important as they advance knowledge on how body gestures contribute to perception/expression of animated agents' emotions. Our research aims to create animated pedagogical agents that can authentically portray personality traits and emotions through lifelike non-verbal cues and speech. This effort seeks to offer a compelling alternative to empathetic human coaches or learning buddies. The study detailed in this paper represents a stride towards reaching this objective.

Acknowledgments. This research is funded by NSF award# 2201019 – "Collaborative Research: Using Artificial Intelligence to Transform Online Video Lectures into Effective and Inclusive Agent-Based Presentations".

References

1. Adamo, N., Mousas, C., Mayer, R.: NSF Award # 2201019 - Collaborative Research: Using Artificial Intelligence to Transform Online Video Lectures into Effective and Inclusive Agent-Based Presentations (2022). https://www.nsf.gov/awardsearch/showAward?AWD_ID=2201019&HistoricalAwards=false&_ga=2.46658222.863696423.1696858525-1636934363.1694122541
2. Annetta, L.A., Holmes, S.: Creating presence and community in a synchronous virtual learning environment using avatars. Intern. J. Instruct. Technol. Dist. Learn. **3**, 27–43 (2006)
3. Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception **33**(6), 717–746 (2004). https://doi.org/10.1068/p5096
4. Bänziger, T., Scherer, K.R.: Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In: Scherer, K.R., Bänziger, T., Roesch, E.B. (eds.) Blueprint for affective computing: A sourcebook, pp. 271–294. Oxford university Press, Oxford, England (2010)

5. Barrett, L.F., Russell, J.A. (eds.): The Psychological Construction of Emotion. The Guilford Press (2015)
6. Baylor, A.L., Kim, S.: Designing nonverbal communication for pedagogical agents: when less is more. Comput. Hum. Behav.. Hum. Behav. **25**(2), 450–457 (2009)
7. Blythe, E., Garrido, L., Longo, M.R.: Emotion is perceived accurately from isolated body parts, especially hands. Cognition **230**, 105260 (2023). https://doi.org/10.1016/j.cognition.2022.105260. Epub 2022 Sep 1. PMID: 36058103
8. Cheng, J., Zhou, W., Lei, X., Adamo, N., Benes, B.: The Effects of Body Gestures and Gender on Viewer's Perception of Animated Pedagogical Agent's Emotions. In: Kurosu, M. (ed.) Human-Computer Interaction. Multimodal and Natural Interaction: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II, pp. 169–186. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-49062-1_11
9. Cui, J., Popescu, V., Adamo-Villani, N., Cook, S.W., Duggan, K.A., Friedman, H.S.: Animation stimuli system for research on instructor gestures in education. IEEE Comput. Graph. Appl.Comput. Graph. Appl. **37**(4), 72–83 (2017). https://doi.org/10.1109/MCG.2017.3271471
10. Ekman, P.:. Emotions revealed. New York/London: Times Books (US)/Weidenfeld & Nicolson (2003)
11. Ekman, P., Friesen, W.: The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica **1**(1), 49–98 (1969)
12. Ennis, C., Hoyet, L., Egges, A., McDonnell, R.: Emotion capture: emotionally expressive characters for games. In: Proceedings of Motion on Games (2013). https://doi.org/10.1145/2522628.2522633
13. Fiorella, L., Mayer, R.E.: Principles based on social cues in multimedia learning: Personalization, voice, embodiment, and image principles. In: Mayer, R.E., Fiorella, L. (eds.) The Cambridge Handbook of Multimedia Learning, 3rd edn, pp. 277–286. Cambridge University Press, New York (in press)
14. Guo, Y.R., Goh, D.H.-L.: Affect in embodied pedagogical agents: meta- analytic review. J. Educ. Comput. Res. **53**(1), 124–149 (2015). https://doi.org/10.1177/0735633115588774
15. Johnson, W.L., Lester, J.C.: Face-to-face interaction with pedagogical agents, twenty years later. Int. J. Artif. Intell. Educ.Artif. Intell. Educ. **26**(1), 25–36 (2016)
16. Karg, M., Samadani, A.A., Gorbet, R., Kühnlenz, K., Hoey, J., Kulić, D.: Body movements for affective expression: a survey of automatic recognition and generation. IEEE Trans. Affect. Comput.Comput. **4**(4), 341–359 (2013)
17. Larsson, P.: Discerning emotion through movement – a study of body language in portraying emotion in animation, pp. 6–7 (2014) MS Thesis. Retrieved from: http://www.divaportal.org/smash/record.jsf?pid=diva2%3A723103&dswid=4060
18. Lawson, A.P., Mayer, R.E., Adamo-Villani, N., Benes, B., Lei, X., Cheng, J.: Do learners recognize and relate to the emotions displayed by virtual instructors? Int. J. Artif. Intell. Educ.Artif. Intell. Educ. **31**, 134–153 (2021)
19. Lhommet, M., Marsella, S.: Expressing emotion through posture and gesture. In: Calvo, R., D'Mello, S., Gratch, J., Kappas, A. (eds.) The Oxford Handbook of Affective Computing, pp. 273–285. Oxford University Press, Oxford (2015)
20. Martha, A.S., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educ. Online, **16**(1) (2019)
21. Mayer, R.E.: Multimedia learning, 3rd edn. Cambridge University Press, New York (2021)
22. Mayer, R.E., Fiorella, L., Stull, A.: Five ways to increase the effectiveness of instructional video. Educ. Tech. Res. Dev. **68**, 837–852 (2020)
23. Mayer, R.E., DaPra, C.S.: An Embodiment effect in computer-based learning with animated pedagogical agents. J. Exp. Psychol. Appl. **18**(3), 239–252 (2012)

24. Mayer, R.E.: Searching for the role of emotions in e-learning. Learn. Instr. **70**, 101213 (2020). https://doi.org/10.1007/978-3-030-90436-4_38
25. Poggiali, J.: Student responses to an animated character in information literacy instruction. Library Hi Tech **36**(1), 29–42 (2018). https://doi.org/10.1108/LHT-12-2016-0149
26. Posner, J., Russell, J.A., Peterson, B.S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol. 2005 Summer **17**(3), 715–734 PMID: 16262989; PMCID: https://doi.org/10.1017/S0954579405050340
27. Rosenberg-Kima, R.B., Baylor, A.L., Plant, E.A., Doerr, C.E.: Interface agents as social models for female students: the effects of agent visual presence and appearance on female students' attitudes and beliefs. Comput. Hum. Behav.. Hum. Behav. **24**, 2741–2756 (2008)
28. Ross, P., Flack, T.: Removing hand form information specifically impairs emotion recognition for fearful and angry body stimuli. Perception **49**(1), 98–112 (2020). https://doi.org/10.1177/0301006619893229
29. Russell, J.A.: Core affect and the psychological construct of emotion. Psychol. Rev. **110**, 145–172 (2003)
30. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
31. Sawada, M., Suda, K., Ishii, M.: Expression of emotions in dance: relation between arm movement characteristics and emotion. Percept. Mot. Skills **97**, 697–708 (2003)
32. Schroeder, N.L., Adesope, O.O., Gilbert, R.B.: How effective are pedagogical agents for learning? a meta-analytic review. J. Educ. Comput. Res. **49**(1), 1–39 (2013)
33. Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: pedagogical agents and learning outcomes. Int. J. Hum. Comput. Stud.Comput. Stud. **66**, 98–112 (2008)