The Effects of Body Gestures and Gender on Viewer's Perception of Animated Pedagogical Agent's Emotions

Justin Cheng, Wenbin Zhou, Xingyu Lei, Nicoletta Adamo(✉), and Bedrich Benes

Purdue University, West Lafayette, IN 47906, USA
nadamovi@purdue.edu

Abstract. The goal of this research is to develop Animated Pedagogical Agents (APA) that can convey clearly perceivable emotions through speech, facial expressions and body gestures. In particular, the two studies reported in the paper investigated the extent to which modifications to the range of movement of 3 beat gestures, e.g., both arms synchronous outward gesture, both arms synchronous forward gesture, and upper body lean, and the agent's gender have significant effects on viewer's perception of the agent's emotion in terms of valence and arousal. For each gesture the range of movement was varied at 2 discrete levels. The stimuli of the studies were two sets of 12-s animation clips generated using fractional factorial designs; in each clip an animated agent who speaks and gestures, gives a lecture segment on binomial probability. 50% of the clips featured a female agent and 50% of the clips featured a male agent. In the first study, which used a within-subject design and metric conjoint analysis, 120 subjects were asked to watch 8 stimuli clips and rank them according to perceived valence and arousal (from highest to lowest). In the second study, which used a between-subject design, 300 participants were assigned to two groups of 150 subjects each. One group watched 8 clips featuring the male agent and one group watched 8 clips featuring the female agent. Each participant was asked to rate perceived valence and arousal for each clip using a 7-point Likert scale. Results from the two studies suggest that the more open and forward the gestures the agent makes, the higher the perceived valence and arousal. Surprisingly, agents who lean their body forward more are not perceived as having higher arousal and valence. Findings also show that female agents' emotions are perceived as having higher arousal and more positive valence that male agents' emotions.

Keywords: Affective pedagogical agents · Body gestures · Gender · Valence · Arousal

1 Introduction

Research has shown that animated pedagogical agents (APA) can be effective in promoting learning [1], but many questions still remain unanswered, particularly concerning their emotional design. With the growing understanding of the complex interplay

M. Kurosu (Ed.): HCII 2020, LNCS 12182, pp. 169–186, 2020.
https://doi.org/10.1007/978-3-030-49062-1_11

between emotions and cognition, there is a need to develop life-like agents that not only provide effective expert guidance, but also convincing emotional interactions with the learner [2–4].

One goal of our research is to develop APAs that can convey clearly perceivable emotions through speech, facial expressions and body gestures. The studies reported in the paper are steps in this direction. They focus on how emotions are conveyed through body cues and, in particular, they examine the extent to which modifications to the range of movement of a set of beat gestures affects viewer' perception of the agent's emotional state. An important issue in bodily expression of emotion research concerns the distinction between a person's encoding of emotion in physical behavior versus an observer's decoding of emotion from observations of the person's behavior. We are concerned with the latter, and whether the observed perceptual effects are moderated by viewer's characteristics such as gender, age, ethnicity and educational level.

There has been considerable debate as to whether posture and movement reliably convey emotions, or rather convey only the intensity of the emotion [5, 6]. We examine how body gestures might convey both the quality of the emotion and its level of activation. We use Russell's [7] model of core affect in which any particular emotion can be placed along two dimensions, valence (ranging from positive to negative), and arousal (ranging from activation to deactivation) and investigate whether and how changes in the motion parameters of a set of body gestures affect the perception of the agent's emotion along both dimensions. We also examine whether the agent's gender has an effect on viewer's perception of the emotional content.

Our studies are important because they may advance not only research on representation of emotion in affective embodied agents, but also psychology research on bodily expression/perception of emotion in general.

2 Background

2.1 Animated Pedagogical Agents

Several studies suggest that the presence of pedagogical agents can improve learning [8, 9]. A meta-analysis by Schroeder showed that lessons with animated pedagogical agents led to statistically significant learning improvements compared to lessons without them [2]. Studies also suggest that APAs could be employed in e-learning environments to enhance users' attitude towards online courses [10]. Agents interacting using multiple modalities appear to lead to greater learning than agents that interact only in a single channel [11, 12].

One reason why pedagogical agents might help to facilitate learning could be that the viewers find them engaging because of their "human-like" personalities [13]. A study by Poggiali showed that students found "animated videos with agents easier to learn from, in part because they held their attention" [14]. The agent's personality helped to contribute towards the student's engagement; study participants reported that the agent's outgoing personality helped them relate to it the same way they relate to an outgoing human instructor. These findings emphasize the importance of establishing a social connection between the agent and the learner. Incorporating emotional design within the pedagogical agents improved learning outcomes within a lesson [15], demonstrating the need for the agents to have recognizable emotions.

Some researchers have investigated the effect of different APA's features on student's learning, engagement, and perception of self-efficacy. Mayer and DaPra [16] examined whether the degree of embodiment of an APA had an effect on students learning of science concepts. Findings showed that students learned better from a fully embodied human-voiced agent that exhibited human-like behaviors and emotions than from an agent who did not communicate using these human-like actions. A study by Gulz and Haake [17] revealed that female students preferred as a learning companion an agent that developed social relationship during the learning activities rather than an agent that was strictly task oriented. Additional empirical studies showed that peer-like agents helped enhance positive affect and motivation for females who learned STEM topics [2, 4]. Other experiments suggest that agent's features such as voice and appearance [18, 19], visual presence [20], non-verbal communication [21] and communication style [22] could impact learning and motivation.

2.2 Expression/Perception of Agents' Emotion and Personality from Body Gestures

The agent's gestures are crucial in conveying its emotional state and personality, as non-verbal cues potentially can make up to 93% of the communication during conversation [23]. A study by Anasingaraju et al. [24] showed that body gestures were the biggest contributors to perceiving the agent's emotion, rather than the character's facial animation and lip sync.

Despite gestures playing a crucial role in conversation, identifying emotions from body gestures alone is not straightforward. Some emotions are easily perceivable from body gestures only, such as anger, sadness, and happiness [25]. However, there are also emotions that are difficult to express with body language alone. Surprise, disgust, and fear are the most difficult emotions to convey from arm movements alone [26]. A study by Atkinson suggests that sadness and disgust were most easily misclassified for each other [27]. A study by Ennis showed that body gestures only without the face caused confusion in differentiating emotions with high arousal [28]. Ennis concluded that body gestures alone caused difficulties in identifying between happy and angry gestures, but in contrast sadness and fear were more identifiable from each other. Karg argues that high valence and low arousal gestures, such as content, were not easy to express from gestures alone [25]. While gestures may help to differentiate between emotions with high and low valence, arousal was more easily identifiable by agent's movement than from still poses.

Several studies that examine agent's expression of personality through body cues can be found in the literature. Compared to an emotion, a personality is defined as a set of permanent or long-lasting complex characteristics that make up how the agent interacts with the environment [29]. In this section we review primarily prior research that compared between extroverted and introverted personalities and that are relevant to our studies. In general, an extroverted agent is likely to express emotions that have positive valence and high arousal, whereas an introverted agent is more likely to show emotions that have negative valence and low arousal.

An agent with an extroverted personality is more likely to show interest and friendliness towards the viewer [30, 31]. As a result, the agent tends to amplify a sense of space [32] by having more horizontal and open arms/hands movements, rotating out the elbows and raising the shoulders. An agent that expresses friendliness, tends to consistently make eye contact with the viewer [32] and stretch the arms towards the viewer [33]. Furthermore, an extroverted character asserts more dominance than an introverted character. This gives the extroverted character an additional incentive to both directly make eye-contact towards the viewer and to amplify a larger sense of space [32].

An agent with an introverted personality tends to show disinterest towards the viewer. As a result, the agent is more likely to focus on minimizing their body size [29]. The upper torso is more likely to stand upright or to lean slightly backwards, whereas the hands are prone to close and touch the agent's own body [32]. When comparing the horizontal spread of the agent's gestures, the introverted agent's horizontal spread is only 10% to 60% of the extroverted agent's horizontal spread [32]. Considering that the introverted agent is more prone to minimizing occupied space, the agent tends to perform fewer out-directing gestures, leading to the introverted character showing more submissiveness [30].

Findings from the studies listed above, as well as best practices in character animation that suggest that high valence/arousal characters are characterized by open and forward body gestures [34] form the basis of the hypotheses of our 2 experiments.

3 Methods

We hypothesize that modifications to the range of movement of 3 beat gestures, e.g., both arms synchronous outward gesture (OG), both arms synchronous forward gesture (FG), and upper body lean (BL), and the agent's gender (G) have significant effects on viewers' perception of the agents' emotion in terms of valence and arousal. More specifically, based on prior research and on best practices in character animation we hypothesize the following:

- Ha(FG): the more forward the gesture, the higher the perceived arousal
- Ha(OG): the more open the gesture, the higher (e.g., more positive) the perceived valence
- Ha(BL): the more forward the body lean, the higher the perceived arousal and perceived valence
- Ha(G): the emotions of female agents are perceived as having higher arousal and higher valence than those of male agents

The 3 beat gestures selected for the study are gestures that are commonly produced by instructors while lecturing and have been shown to convey some information about the speaker's emotional state, personality or status [35]. For each gesture the range of movement is varied at 2 discrete levels. Table 1 lists the factors and levels used in the studies.

Two experiments were conducted; one study used a within-subjects design and metric conjoint analysis; the other study used a between-subject design and linear regression.

Table 1. Factors and levels

Factor	Level 1	Level 2
OG	Arms/hands are close to body on the sides	Arms/hands are spread apart horizontally
FG	Hands are right in front of the body	Arms/hands are stretched in front of the agent
BL	Body leans backwards	Body leans forwards
G	Agent is male	Agent is female

The stimuli for both studies were sets of 12-s animation clips generated using partial factorial designs; each clip showed a different combination of body gestures and ranges of motion. 120 subjects participated in the first study and 300 subjects participated in the second one. In the first study subjects were asked to watch the stimuli clips and rank them from highest to lowest arousal and valence. In the second study subjects were divided into two groups: one group was assigned the clips featuring the female agent and one group the clips featuring the male agent. All subjects were asked to watch the stimuli clips and rate the valence and arousal of each clip using a 7-point Likert scale.

The stimuli in both studies were two sets of 12-s animation clips; in each clip an animated agent who speaks and gestures, gave a lecture segment on binomial probability. All animation clips were assembled within the Unity game engine and the gesture animations and slide timings were manually synced using the Unity's Timeline feature. The agents were framed from thigh-up, at a ¾ view towards the viewer so that FG and BL levels were clearly visible.

Both agents are commercially available 3D character rigs whose joint structure was modified in order to be compatible with Unity's character animator feature. The agents' gestures were motion captured and manually blended together; the agents' lip-sync animations were generated with a Unity script. Camera angle, background, lighting, and speech were kept the same in every clip. The agent's faces were blurred out to eliminate potential confounding variables.

The full quote that the agents spoke was the following: "A success is defined by you as one or more of the possible outcomes. For example, a success of rolling a die could be that you rolled a number greater than four." While the agent was speaking the first sentence, the agent would point towards the viewer with his or her left hand. This gesture was not modified and remained the same in all animation clips. During the second sentence, the agent would smoothly transition to another gesture, modified by the offset script with parameters that varied in each clip, which involved the agent opening out his or her hands and leaning his or her body. The agent returned to a neutral standing position at the end of each gesture.

The labelling format for the clips was in the form of [G]_OG[x]-FG[x]-BL[x], where [G] was replaced by the gender initial of M or F, and [x] was replaced by 1 or 2 to indicate level.

4 Study 1

4.1 Design

A metric conjoint analysis and a partial factorial design were used for the first study. As the study included 4 different factors (the 3 gestures + agent's gender) each varied at 2 levels, a full factorial design would have required 16 different clips. Such number of animations would have been difficult to rank. Hence, 8 representative combinations of the factors were created according to a fractional factorial design generated with JMP PRO 14 statistical software. This design allowed us to examine main effects as well as 2-factor interactions. The 8 representative combinations of the factors and levels are listed in Table 2. Figure 1 shows frames extracted from the 8 animation clips used in the first study.

The evaluation instrument was an online survey designed using Qualtrics software. In the survey, subjects were presented with 2 ranking tasks: (1) rank the 8 clips in terms of arousal and (2) rank the 8 clips in terms of valence from highest to lowest. For each ranking task, all eight clips were simultaneously loaded, with the clip order randomized. Participants needed to drag and drop the clips in order to rank them, with the clips that had highest perceived valence or arousal placed at the top of the page. Demographic information such as the participant's age, gender, and highest level of completed education were also collected. The survey's time limit was set to 20 min. The study used simple random sampling and anyone over the age of 18 was eligible to participate. 120 subjects were recruited through Amazon Mechanical Turk but only 103 responses were considered. The subjects' demographics were the following: male (66), female (37); 18–30 years old (45), 30+ years old (57); no-degree (32) and degree (71).

A 4-way ANOVA test and a linear regression model were applied to test for main effects and two-way interactions. A Chi-squared test was applied to test the independence of the valence and arousal rankings. The study assumed that the data distribution was normal, and all ranking entries were independent.

Table 2. Factors and levels combinations

#	Label	OG	FG	BL	G
1	M_OG2-FG1-BL2	2	1	2	M
2	M_OG1-FG1-BL1	1	1	1	M
3	M_OG1-FG2-BL2	1	2	2	M
4	M_OG2-FG2-BL2	2	2	1	M
5	F_OG1-FG1-BL2	1	1	2	F
6	F_OG2-FG2-BL2	2	2	2	F
7	F_OG1-FG2-BL1	1	2	1	F
8	F_OG2-FG1-BL1	2	1	1	F

Fig. 1. Frames extracted from each of the 8 stimuli clips used in the first study

4.2 Results

A total of 120 study participants completed the survey, however only 103 responses were considered. 17 responses were discarded because completion times were below our 22 s time filter. Table 3 shows the average mean rankings and standard deviations, and Table 4 shows the statistical analysis results (p for p-value and β for linear regression model coefficient). A higher mean rank meant that clip had lower perceived valence or arousal.

Table 3. Primary study full results mean rankings

Clip	Valence M	Valence SD	Arousal M	Arousal SD
M_OG1-FG1-BL1	5.05	2.16	4.32	2.2
M_OG1-FG1-BL1	4.98	2.09	4.43	2.29
M_OG1-FG2-BL2	5.03	2.2	3.97	2.29
M_OG2-FG2-BL1	5.04	2.36	4.5	2.33
F_OG1-FG1-BL2	4.27	2.47	4.59	2.36
F_OG2-FG2-BL2	3.77	2.23	4.93	2.2
F_OG1-FG2-BL1	4.03	2.27	4.69	2.23
F_OG2-FG1-BL1	3.83	2	4.56	2.32

Gender (G) was a significant main effect for perceived valence ($p =$ nearly 0.000) and for perceived arousal ($p = 0.015$). OG was close to statistical significance, ($p = 0.083$) for perceived valence. According to the linear regression model for valence and

Table 4. Primary study full results statistical analysis test

Factor	Valence p	Valence β	Arousal p	Arousal β
OG	0.319	−0.156	0.083	−0.276
FG	0.663	−0.068	0.808	−0.038
BL	0.709	0.058	0.927	−0.014
G	0	−1.048	0.015	−0.388
OGxFG	0.852	0.03	0.447	−0.122
OGxBL	0.575	−0.088	0.273	−0.174
FGxBL	0.213	−0.194	0.784	−0.044

arousal, changing G from male to female changed the mean valence ranking by nearly a full rank ($\beta = -1.048$), and about four-tenths of an arousal rank ($\beta = -0.388$). Overall, the data supported that gender was a significant main effect for both valence and arousal rankings; the clips featuring the female agent were ranked significantly higher for valence and arousal than the clips featuring the male agent.

Table 5. P values for perceived valence for demographic subsets – study 1

	Age subset		Gender subset		Education level subset	
Factor	18–30	30+	Female	Male	Degree	No-degree
OG	0.034	0.675	0.166	0.845	0.281	0.864
FG	0.449	0.967	0.166	0.612	0.710	0.820
BL	0.042	0.181	0.878	0.725	0.528	0.776
G	0.000	0.000	0.001	0.000	0.000	0.000
OGxFG	0.421	0.530	0.538	0.482	0.911	0.608
OGxBL	0.570	0.209	0.538	0.242	0.656	0.088
FGxBL	0.603	0.225	0.412	0.349	0.081	0.690

In regard to the demographic subsets, Table 5 and Table 6 report the results of the statistical analyses for perceived valence and arousal respectively. For the valence rankings OG, BL, and G were significant main effects for the 18–30 years old group. OG and BL had β values of −0.5 and 0.478 respectively. For the 30+ years old subset for valence, only G was a significant main effect; no significant main effects were found for arousal.

In the male participants subset, only G was a significant main effect for valence. In the female participants group, G was significant for valence, as well as for arousal.

In regard to the 'no degree' subset only G was a significant main effect for valence; no significance was found for arousal. In the degree subset, G was a significant main effect for valence; no significance was found for arousal.

Table 6. P values for perceived arousal for demographic subset – study 1

	Age subset		Gender subset		Education level subset	
Factor	18–30	30+	Female	Male	Degree	No-degree
OG	0.153	0.166	0.244	0.193	0.256	0.157
FG	0.381	0.596	0.685	1.000	0.883	0.513
BL	0.213	0.271	0.761	0.733	0.420	0.174
G	0.019	0.254	0.023	0.185	0.062	0.115
OGxFG	0.356	0.775	0.648	0.544	0.256	0.744
OGxBL	0.644	0.308	0.960	0.161	0.085	0.549
FGxBL	1.000	0.744	0.418	0.791	0.304	0.301

The reported R-squared values for both the full dataset and all of the subsets had a minimum of 0.044 (Degree) to a maximum of 0.115 (No Degree) for valence, and 0.011 (Male) to 0.04 (18–30 y/o). As for the Chi-square test, significance was found in both the full dataset ($p = 0.037$) and the male subset ($p = 0.011$).

4.3 Discussion

Results of the statistical analyses support our hypothesis that gender is a significant main effect across all subjects and for each demographic subset: the clips featuring the female agent were ranked significantly higher for valence and arousal than the clips featuring the male agent.

When examining different demographic subsets, results show that for the 18–30 years old group OG and BL were significant main effects for valence rankings, with OG raising the valence ranking ($\beta = -0.5$) and BL decreasing valence ranking ($\beta = 0.478$) when the factor levels changed from level 1 to level 2 for both factors. This suggests that for the 18–30 years old group increasing the agent's OG from level 1 to level 2 (e.g., extending the agent's arms/hands outward more) results in higher perceived valence. This finding supports our OG hypothesis. However, leaning the agent's body forward decreases perceived valence; this finding does not support our BL hypothesis.

As for arousal rankings, only the 18–30 years old subset and the female subset suggests that G influences the arousal rankings ($P = 0.019$ and 0.023). According to the linear regression model, modifying G from male to female for the listed groups would result in higher perceived arousal (β values of -0.566 and -0.608).

The Chi-squared test that was performed on the main dataset, yielded a p-value of 0.037, suggesting that the valence and arousal rankings were not-independent of each other for the full dataset. This was also shown for the male subset, where the p-value was 0.011.

It should be noted that for all the datasets, the reported R-squared values were very low. From the main dataset, the valence R2 value was 0.056 and the arousal R2 value was 0.013. The R2 values from the subsets all did not exceed 0.150. The low R2 suggests that even if the data points suggest a relationship between factors and perceived valence/arousal, the exact relationship may differ from the suggested linear model.

5 Study 2

5.1 Design

After a careful review of the findings from study 1, we identified two potential issues with the experiment design. First, we realized that the agent design, e.g., appearance and voice, may have introduced gender comparison biases. In other words, subjects might have ranked the clips featuring the female agent higher in valence and arousal because of the intrinsic characteristics of the female agent design and because of the quality of her voice, rather than because of the mere difference in gender. Second, subjects commented that the ranking tasks were quite difficult to perform and rating the clips in terms of perceived valence and arousal would have been easier than ranking them. These issues prompted us to conduct a follow- up experiment; we refer to this second experiment as study 2.

Table 7. Study 2 clip combinations

#	Label	OG	FG	BL	G
1	M_OG2-FG1-BL2	2	1	2	M
2	M_OG1-FG1-BL2	1	1	2	M
3	M_OG2-FG2-BL2	2	2	2	M
4	M_OG1-FG2-BL2	1	2	2	M
5	M_OG1-FG1-BL1	1	1	1	M
6	M_OG2-FG2-BL1	2	2	1	M
7	M_OG1-FG2-BL1	1	2	1	M
8	M_OG2-FG1-BL1	2	1	1	M
9	F_OG2-FG1-BL2	2	1	2	F
10	F_OG1-FG1-BL2	1	1	2	F
11	F_OG2-FG2-BL2	2	2	2	F
12	F_OG1-FG2-BL2	1	2	2	F
13	F_OG1-FG1-BL1	1	1	1	F
14	F_OG2-FG2-BL1	2	2	1	F
15	F_OG1-FG2-BL1	1	2	1	F
16	F_OG2-FG1-BL1	2	1	1	F

Study 2 used a between-subjects design with 300 participants; the stimuli were 16 12-s animation clips. Subjects were recruited through Amazon Mechanical Turk and divided into 2 groups of 150 participants each: one group viewed only the 8 clips featuring the male agent and one group viewed only the 8 clips featuring the female agent. Study 2 used a partial factorial design generated using JMP Pro 14 statistical software. Table 7 lists the clip combinations and Fig. 2 shows frames extracted from each one of the 16 stimuli clips.

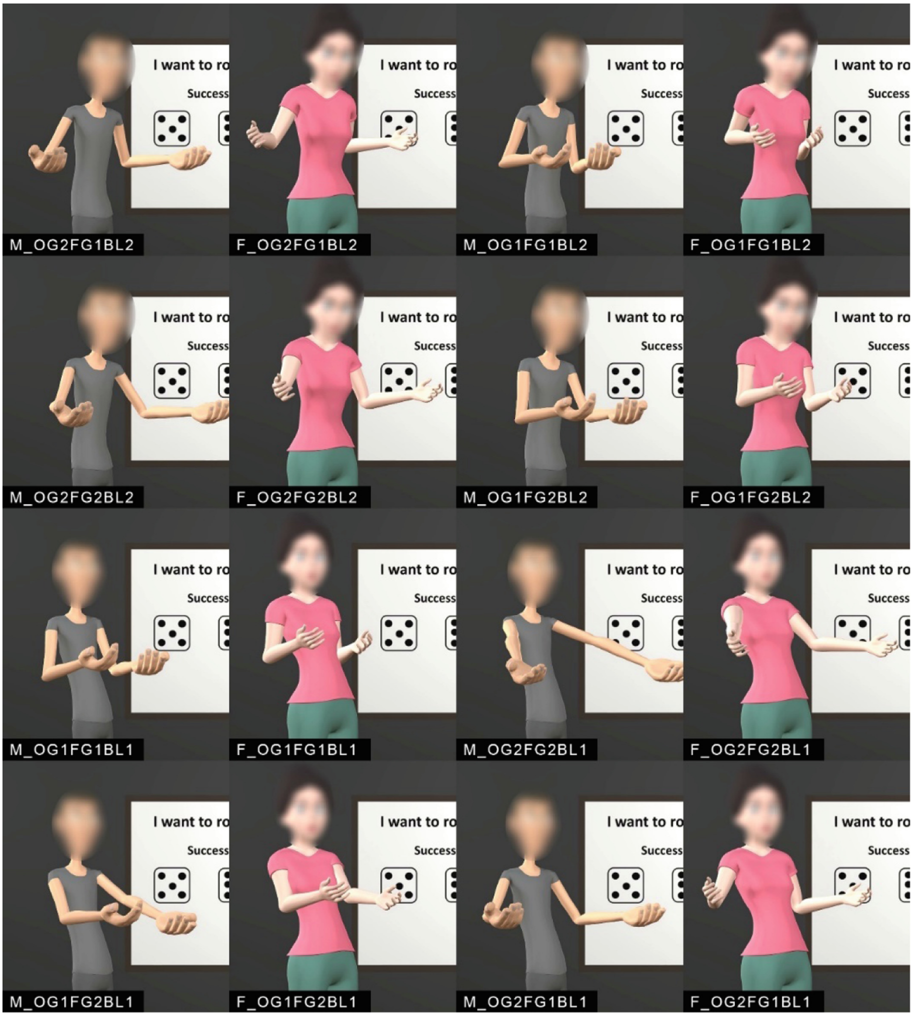

Fig. 2. Frames extracted from the 16 stimuli clips used in study 2

The evaluation instrument was an online survey designed using Qualtrics software. In the survey, subjects in each group were asked to watch the 8 video clips (presented in randomized order) and rate each clip for valence and arousal using a 7-pointLlikert scale (1 = low arousal; 7 = high arousal; 1 = highly negative valence; 7 = highly positive valence).

The same set of statistical tests and hypotheses as in study 1 were used. In study 2, a higher number (rating) meant that the clip had higher perceived valence/arousal.

5.2 Results

A total of 300 subjects completed the survey; the subjects' demographics were the following: male (182), female (89), 18–30 years old (53), 30+ years old (217), no degree (83), degree (186). Responses with "prefer not to answer" were not considered in the subsets' analyses. Out of the 300 collected responses, 27 were discarded, as the completion times were below our predesignated time filter of 11 s. The study assumed that the data distribution was normal, and all ratings were independent. Table 8 shows the average mean rankings (higher M means higher rated), and Table 9 shows the statistical analysis results.

Results of a 4-way ANOVA test performed across all subjects showed that all main factors were significant for perceived valence (OG $P = 0.003$, FG $P = 0.012$, BL $P = 0.022$, G $P =$ nearly 0.000). As for perceived arousal, only OG and FG were significant main factors (OG $P = 0.000$, FG $P = 0.017$).

Table 10 and Table 11 show the results of the statistical analyses for the demographic subsets for perceived valence and arousal, respectively. In the 18–30 years old subset OG, FG, and BL were found to be significant main factors for perceived valence, and OG and BL were significant main factors for arousal. In the 30+ years old subset OG,

Table 8. Mean and standard deviation values for study 2

Clip	Valence M	Valence SD	Arousal M	Arousal SD
M_OG2-FG1-BL2	45.64	10.02	49.85	11.14
M_OG1-FG1-BL2	45.27	9.86	49.54	11.09
M_OG2-FG2-BL2	46.3	10.19	50.31	10.19
M_OG1-FG2-BL2	45.61	9.94	48.03	12.36
M_OG1-FG1-BL1	45.24	10.23	48.15	10.86
M_OG2-FG2-BL1	47.53	10.7	52.27	10.31
M_OG1-FG2-BL1	46.48	9.53	49.82	11.51
M_OG2-FG1-BL1	46.62	10.15	50.32	11.09
F_OG2-FG1-BL2	47.76	10.2	49.22	10.54
F_OG1-FG1-BL2	47.04	10.69	48.11	11.86
F_OG2-FG2-BL2	49.53	9.97	51.77	10.76
F_OG1-FG2-BL2	47.83	9.56	48.35	11.61
F_OG1-FG1-BL1	47.04	9.88	47.67	11.34
F_OG2-FG2-BL1	50.6	9.49	51.76	11.21
F_OG1-FG2-BL1	49.32	8.99	50.32	10.82
F_OG2-FG1-BL1	49.97	10.2	50.76	10.6

Table 9. p and β values for study 2

Factor	Valence p	Valence β	Arousal p	Arousal β
OG	0.003	1.272	0	2.038
FG	0.012	1.082	0.017	1.136
BL	0.022	−0.982	0.12	−0.74
G	0	2.548	0.932	−0.04
OGxFG	0.839	−0.086	0.451	0.358
OGxBL	0.355	−0.396	0.601	−0.25
FGxBL	0.667	−0.184	0.152	−0.684

FG, and G were significant main factors for perceived valence and OG and FG were significant main factors for perceived arousal.

Table 10. P values for perceived valence for demographic subsets – study 2

	Age subset		Gender subset		Education level subset	
Factor	18–30	30+	Female	Male	Degree	No-degree
OG	0.080	0.014	0.013	0.037	0.039	0.010
FG	0.055	0.063	0.322	0.014	0.041	0.108
BL	0.001	0.337	0.079	0.144	0.136	0.041
G	0.100	0.000	0.000	0.024	0.000	0.045
OGxFG	0.255	0.360	0.991	0.672	0.958	0.528
OGxBL	0.675	0.192	0.930	0.250	0.401	0.609
FGxBL	0.463	0.360	0.835	0.699	0.878	0.580

For gender, in the male subset OG, FG, and G were significant main factors for perceived valence, whereas OG and FG were significant for perceived arousal. In the female subset OG and G were significant main factors for perceived valence, while only OG was significant for arousal. G had a surprising β of 5.084.

In regard to the level of education subsets, in the no degree subset OG, BL and G were significant main factors for perceived valence, while OG and FG were significant for perceived arousal. In the degree subset OG, FG and G were significant main factors for perceived valence, while OG only was significant for arousal.

The reported R-squared values for both the full dataset and all of the subsets had a minimum of 0.013 (Male) to the maximum of 0.073 (Female) for valence, and 0.013 (Full and Degree) to 0.33 (18–30 y/o).

Table 11. P values for perceived arousal for demographic subsets – study 2

	Age subset		Gender subset		Education level subset	
Factor	18–30	30+	Female	Male	Degree	No-degree
OG	0.082	0.000	0.004	0.001	0.001	0.006
FG	0.364	0.019	0.215	0.030	0.093	0.004
BL	0.008	0.836	0.094	0.504	0.075	0.965
G	0.246	0.859	0.035	0.058	0.836	0.577
OGxFG	0.193	0.991	0.304	0.950	0.622	0.654
OGxBL	0.823	0.408	0.672	0.416	0.506	0.977
FGxBL	0.503	0.208	0.603	0.175	0.355	0.201

5.3 Discussion

Findings from the statistical analyses show that all main factors were significant for perceived valence; OG and FG were significant main factors for perceived arousal. Results support 3 of our hypotheses, e.g., OG, FG, and G affect perceived valence, and OG and FG affect perceived arousal. After applying the linear regression model to the full dataset, the model still shows that G is the most influential factor for perceived valence ($\beta = 2.548$) and perceived valence increases when G is changed from male to female. OG and FG increase the valence rating by a smaller amount when the levels are changed from 1 to 2, with OG $\beta = 1.272$ and FG $\beta = 1.082$. As for arousal, OG is the most influential factor when increased from level 1 to level 2 ($\beta = 2.038$). FG also increases perceived arousal when changed from level 1 to 2 ($\beta = 1.136$). Although BL was found to be a significant main effect, BL's $\beta = -0.982$, hence perceived valence decreases when BL changes from level 1 to level 2. This finding does not support our BL hypothesis.

The same significant main effects were found for the different data subsets with a few exceptions. All subgroups except the 18–30 years old group demonstrated that OG is a significant factor for influencing perceived valence and perceived arousal. FG was a significant factor for the male subset and the degree subset for valence and was a significant main effect for the 30+ years old subset and the no degree subset for arousal. BL was only significant for the 18–30 years old subset and the no degree subset for valence, and the 18–30 years old subset for arousal. G was a significant factor for the 30+ years old subset, female subset, no degree subset, and degree subset for valence, and a significant main factor for arousal for the female subset.

The linear regression models performed on the subsets in study 2 aligned with the models performed in study 1 for many of the factors. Two notable exceptions were BL for valence ratings for the 18–30 years old subset, and G for valence rating for the female subset. According to the regression models, modifying BL from level 1 to level 2 would result in $\beta = -3.34$ in comparison to the first study BL's $\beta = -0.982$. As for the female subset, modifying G from male to female results in $\beta = 5.084$ compared to the first study G's $\beta = 2.548$.

Overall, the R-squared values were low. The full dataset's R^2 was 0.025 for valence and 0.013 for arousal. The largest R^2 value when looking at the data subset was 0.073, which was still very low. Just like Study 1, the low R^2 suggests that even if the data points suggest a relationship between factors and perceived valence/arousal, the exact relationship may differ from the suggested linear model.

6 Conclusion and Future Work

The findings from the two studies are consistent in part with results of prior experiments reported in the literature review. They provide evidence that a positive-valence, high arousal pedagogical agent tends to display open body gestures, e.g., he/she tends to open the arms and hands more outwards, as in study 2 OG was tested to be significant in affecting both perceived valence and arousal. The linear regression models showed that there was a positive relationship between the arms opening out and the viewer's perceived valence and arousal. Both the female and male agent's OG2-FG2-BL1 clips in study 2 were perceived as having either the highest valence or arousal, the arms were forward and spread out in both those clips. Findings also provided evidence that a positive-valence, high-arousal pedagogical agent tends to reach out to the viewer more with arms and hands, as FG was a statistically significant main effect for both perceived valence and arousal in study 2. The linear regression models showed a positive relationship between stretching the arms forward and the perceived valence and arousal. The clips with the highest perceived valence or arousal were also the OG2-FG2-BL1 clips as mentioned above.

The studies did not support prior research findings according to which highly active/engaged pedagogical agents would tend to lean their body forward more. While BL was tested to be significant in study 2 (both valence and arousal from the 18/30+ years old subset and just valence from the full dataset), the linear regression model suggested that leaning the body forward decreased the viewer's perceived valence and arousal of the agent. One possible explanation is that modifying BL via the Unity custom script would modify the agent's body lean independently from the hand locations. For example, if BL was adjusted backward, the hands and arms would not be dragged with the body too. As a result, the backwards leaning motion of the agent could have been perceived more as a balancing action than an intentional action of the agent trying to reach out towards the viewer. This was further demonstrated in study 2 as the clips with the highest mean valence and arousal rating were both M_OG2-FG2-BL1 and F_OG2-FG2-BL1 clips; in these clips the agent's hands were spread wide open and towards the viewer while the agent's body was leaned backwards.

Gender was the most significant main factor for both perceived valence and arousal in both studies. Study 1 showed that G was the most significant factor for valence rankings. G was also shown to be significant in study 2 for valence, with the linear regression model stating that G influenced the valence and arousal ratings more than any of the body gesture factors.

The results of our studies in part support findings from prior studies which suggest that female agents are perceived as more supportive/positive. The female clips had the highest perceived mean valence rankings or ratings in both studies. As for male agents being perceived to be more engaging/active (high arousal) [36], both studies' findings were not able to statistically support this claim, despite the fact that the clips with the highest perceived arousal were all male clips. Prior findings that suggest that participants find agents of opposite gender more engaging were not supported by our studies. Study 2 suggested that for the female subset G was a significant main effect and female clips had significant higher perceived arousal.

While the studies also attempted to determine if two-way factor interactions OGxFG, OGxBL, and FGxBL were significant, data analyses from both the first and second studies were unable to confirm if these interactions influenced valence or arousal perception. Only main effects OG, FG, BL and G had significance.

Our studies focused on body poses only, other important factors such as gesture frequency and movement speed were not considered. In future experiments we will examine how these variables affect perception of the agent's emotions. In addition, in future work we will also investigate which specific visual features of an agent and degree of embodiment and personalization influence perception of the agent and for which types of learners.

Our research aims to develop APAs that can display believable and convincing personality and emotions through life-like gestures, speech, body movements, and facial expressions, and hence be an effective alternative to expert and caring human tutors or learning companions. The main goal is to generate evidence-based guidelines for how to incorporate affective (or emotional) features in APAs, in order to improve learning. The studies reported in the paper are first steps in this direction.

Acknowledgements. The work reported in the paper is supported in part by NSF – Cyberlearning Collaborative Research: Multimodal Affective Pedagogical Agents for Different Types of Learners, Award Number: 1821894. We thank Purdue Statistical Consulting Services for their help with the statistical analyses.

References

1. Schroeder, N., Adesope, O.O., Barouch Gilbert, R.: How effective are pedagogical agents for learning? A meta-analytic review. J. Educ. Comput. Res. **49**(1), 1–39 (2013)
2. Kim, Y., Baylor, A.L.: Pedagogical agents as social models to influence learner attitudes. Educ. Technol. **47**(01), 23–28 (2007)
3. Zhou, W., Cheng, J., Lei, X., Benes, B., Adamo, N.: Deep learning-based emotion recognition from real-time videos. In: Proceedings of HCI International 2020, Copenhagen, Denmark (2020, in press)
4. Kim, Y., Lim, J.: Gendered socialization with an embodied agent: creating a social and affable mathematics learning environment for middle-grade females. J. Educ. Psychol. **105**(4), 1164–1174 (2013)
5. Ekman, P., Friesen, W.: The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica **1**(1), 49–98 (1969)
6. 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)
7. Russell, J.A.: Core affect and the psychological construct of emotion. Psychol. Rev. **110**, 145–172 (2003)

8. Johnson, W.L., Lester, J.C.: Face-to-face interaction with pedagogical agents, twenty years later. Int. J. Artif. Intell. Educ. **26**(1), 25–36 (2016)
9. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educ. Online, **16**(1) (2019)
10. Annetta, L.A., Holmes, S.: Creating presence and community in a synchronous virtual learning environment using avatars. Int. J. Instr. Technol. Distance Learn. **3**, 27–43 (2006)
11. Alseid, M., Rigas, D.: Three different modes of avatars as virtual lecturers in elearning interfaces: a comparative usability study. Open Virtual Reality J. **2**, 8–17 (2010)
12. Lusk, M.M., Atkinson, R.K.: Varying a pedagogical agent's degree of embodiment under two visual search conditions. Appl. Cogn. Psychol. **21**, 747–764 (2007)
13. Dehn, D.M., Van Mulken, S.: Impact of animated interface agents: a review of empirical research. Int. J. Hum. Comput. Stud. (2000). https://doi.org/10.1006/ijhc.1999.0325
14. Poggiali, J.: Student responses to an animated character in information literacy instruction. Library Hi Tech **36** (2018). https://doi.org/10.1108/lht-12-2016-0149
15. Mayer, R.E., Estrella, G.: Benefits of emotional design in multimedia instruction. Learn. Instr. **33**, 12–18 (2014)
16. 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)
17. Gulz, A., Haake, M.: Social and visual style in virtual pedagogical agents. In: Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors, 10th International Conference on User Modelling (2005)
18. Domagk, S.: Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. J. Media Psychol. **22**(2), 84–97 (2010)
19. Mayer, R.E.: Principles based on social cues in multimedia learning: personalization, voice, image, and embodiment principles. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning, 2nd edn, pp. 345–368. Cambridge University Press, New York (2014)
20. 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. **24**, 2741–2756 (2008)
21. Baylor, A.L., Kim, S.: Designing nonverbal communication for pedagogical agents: when less is more. Comput. Hum. Behav. **25**(2), 450–457 (2009)
22. 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. **66**, 98–112 (2008)
23. Larsson, P.: Discerning emotion through movement – a study of body language in portraying emotion in animation, pp. 6–7, May 2014. http://www.diva-portal.org/smash/get/diva2:723103/FULLTEXT01.pdf
24. Anasingaraju, S., Popescu, V., Adamo, N., Wu, M.L.: Digital learning activities delivered by eloquent instructor avatars: scaling with problem instance. In: Proceedings of SIGGRAPH Asia 2016 – Education Symposium. ACM Digital Library (2016)
25. 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. **4**(4), 341–359 (2013)
26. 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)
27. 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
28. 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

29. André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T.: Exploiting models of personality and emotions to control the behavior of animated interactive agents. In: Workshop on "Achieving Human-Like Behavior in Interactive Animated Agents" in Conjunction with the Fourth International Conference on Autonomous Agents, pp. 3–7 (2000)
30. Allbeck, J., Badler, N.: Toward representing agent behaviors modified by personality and emotion. In: Embodied Conversational Agents at AAMAS, vol. 2, pp. 15–19 (2002). https://doi.org/10.1.1.19.2054
31. Mehrabian, A.: Analysis of the big-five personality factors in terms of the PAD temperament model. Aust. J. Psychol. (1996). https://doi.org/10.1080/00049539608259510
32. 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
33. Ball, G., Breese, J.: Relating personality and behavior: posture and gestures, pp. 196–203 (2006). https://doi.org/10.1007/10720296_14
34. Roberts, S.: Character Animation in 3D. Focal Press, Oxford (2004)
35. Cui, J., Adamo, N., Popescu, V.: Charismatic and eloquent instructor avatars with scriptable gesture. In: Proceedings of SIGGRAPH 2014 - Talks, Vancouver, August 2014 (2014)
36. Kim, Y.: Pedagogical agents as learning companions: the effects of agent affect and gender on learning, interest, self-efficacy, and agent persona. Ph.D. thesis, FSU (2003). https://diginole.lib.fsu.edu/islandora/object/fsu:181467/datastream/PDF/view