The Visual Computer
https://doi.org/10.1007/s00371-024-03622-w

RESEARCH

ASAP: animation system for agent-based presentations

Minsoo Choi[1] · Christos Mousas[1] · Nicoletta Adamo[1] · Sanjeevani Patankar[1] · Klay Hauser[1] · Fangzheng Zhao[2] · Richard E. Mayer[2]

Accepted: 27 August 2024

Abstract
We introduce an animation system that transforms instructional videos into 3D animations with pedagogical agents, supplemented by animation editing tools to foster expressive, agent-based presentations, thereby enhancing educational benefits. Our system extracts the lecturer's motion from the imported instructional video. Once the data extraction is complete, the system retargets this motion to the virtual agent and integrates it into a virtual classroom environment. Subsequently, it provides a GUI-based animation editing tool, offering a range of resources, such as motion assets (e.g., upper body gestures, facial expressions), which enable users to layer them on top of the extracted motion to make the pedagogical agent's motions more engaging. To evaluate our system, we conducted a user study encompassing non-expert and expert groups, employing a mixed-method approach by collecting both quantitative and qualitative data. The results demonstrated our system's educational value, functionality, and usability. Furthermore, the comparative analysis between the non-expert (people with no animation experience) and expert (people with prior animation experience) user groups provided distinct perspectives on our system, reflecting differences in the user's animation experience and expertise. However, both groups reported similar usability and task load levels, indicating that non-experts can use our system efficiently to produce expressive agent-based presentations. We plan to release our system as an open source, cross-platform solution to help educators create engaging agent-based presentations.

Keywords Character animation · Motion editing · Pedagogical agents · Human-computer interaction

1 Introduction

Kentnor [25] stated, "Online education is no longer a trend. Rather, it is mainstream." Online lecture videos have become a popular multimedia learning content, and the number of students participating in online lectures has increased [27]. Researchers have conducted studies in multimedia learning to enhance educational benefits. Mayer [37] presented the cognitive theory of multimedia learning and emphasized fostering generative processing, such as integrating human-like gestures, to reduce the required cognitive load from learners. Lawson et al. [30] pointed out the importance of motivation in multimedia learning. Furthermore, Pekrun and Stephen [53] stated that the learner's emotional state affects motivation, and Loderer et al. [33] addressed the effect of emotion on academic learning. Other studies [21, 31] indicated the effectiveness of displaying positive-active emotions from instructors regarding learning outcomes. As these previous studies presented, multimedia learning necessitates numerous factors to enhance educational benefits.

✉ Minsoo Choi
choi714@purdue.edu

Christos Mousas
cmousas@purdue.edu

Nicoletta Adamo
nadamovi@purdue.edu

Sanjeevani Patankar
spatank@purdue.edu

Klay Hauser
hauser12@purdue.edu

Fangzheng Zhao
fangzheng.zhao@psych.ucsb.edu

Richard E. Mayer
mayer@ucsb.edu

[1] Purdue University, West Lafayette, IN 47907, USA

[2] University of California Santa Barbara, Santa Barbara, CA 93106, USA

Published online: 03 October 2024

However, it is challenging to say that all instructional lecture videos fulfilled these theories and findings, and this limitation leads to educational losses. For example, if an instructor in a video lecture delivered a lecture without expressive non-verbal behaviors, students would not be exposed to a "good" learning experience [64]. Researchers have reported that the high cost of instructional videos is an issue [15]. For example, Nikopoulou-Smyrni and Nikopoulos [48] described the production of lecture videos as a time-consuming task, and Rubenstein [58] reported substantial costs of good online course design.

To overcome these issues, we implemented an animation system that converts instructional videos into 3D animations with pedagogical agents and provides animation editing tools to its users to make the pedagogical agents-based presentation more engaging. Our system consists of two main steps: data extraction and animation enhancement. In the data extraction step, our system extracts the lecturer's pose from the imported video lecture and converts the sequence of these poses into an animation clip to serve as baseline motion. Subsequently, the system proceeds to the animation enhancement step. In this step, our system retargets the animation clip to the pedagogical agents and integrates it into a virtual classroom environment. It then offers a GUI-based animation editing tool to enable users to 1) edit the baseline motion by adding beat, deictic, and affective body gestures and facial expressions, 2) change the pedagogical agent, and 3) add slides as lecture resources.

To evaluate our system's usability, functionality, and potential educational value, we recruited a non-expert group ($N = 8$) and an expert group ($N = 8$) in computer graphics and animation. In our study, we asked participants to use the system to transform the video lecture into an expressive agent-based presentation and modify the pedagogical agent to enhance its expressiveness. Upon completing the task, we conducted a mixed-method study. First, our participants completed two questionnaires— the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX). Second, they participated in an interview session, responding to open-ended questions. Based on our study and collected data, we aimed to understand how inexperienced and experienced users interacted with our system and whether they evaluated it as an effective and easy-to-use system for converting instructional videos into agent-based presentations.

We organized this paper as follows. In Sect. 2, we discuss the work related to our project. We present the details of implementation in Sect. 3 and the experimental design and measurement in Sect. 4. In Sect. 5, we report the result of the experiment. We discuss the reported results in Sect. 6 and the limitations in Sect. 7. Finally, in Sect. 8, we conclude and discuss future work.

2 Related works

2.1 Computational Character Animation Synthesis

Researchers have explored data-driven methods to synthesize animation for virtual characters using existing motion data [7, 50, 52]. Researchers have used numerous techniques in character animation synthesis, such as warping, blending, or interpolation. Lui and Zhang [34] introduced a methodology based on the relative spacetime transformations between different identify-independent coordinates to synthesize the stylized motion. Jovane et al. [24] presented a motion warping method that updates the animations based on the visual motion features. In the context of blending, Kovar and Gleicher [28] devised a registration curve that eliminates the necessity for manual input within the blending process, thereby facilitating automatic interpolation. Neff and Kim [47] employed blending motions by utilizing dynamic motion warping to synthesize stylistic animation. Regarding interpolation, Mukai and Kuriyama [44] proposed a method to predict missing motion data from a sample motion.

Researchers have used other types of data for character animation synthesis and editing. Some researchers focus on text data to synthesize and edit the character motion. Mousas and Anagnostopoulos [40] presented a character animation environment based on predefined commands. Oshita [51] introduced an animation system that used script-like texts, such as movie scripts, to generate animations. Kim et al. [26] introduced FLAME, a text-based motion generation model. Zhang et al. [70] developed a system that allows users to create, edit, preview, and render animation using text descriptions. Researchers have also explored audio sources as data for character animation. Sauer and Yang [59] introduced a system synthesizing character animations from extracted musical features, and Alexanderson et al. [2] presented a diffusion model to synthesize human motion driven by audio. Moreover, Cardle et al. [12] analyzed the music and created the curves based on the musical features, such as the beat. By blending the motion curve with these created curves, their system guided users to synchronize animations with the music.

Researchers have considered character animation synthesis and editing as optimization problems to provide realistic character animation. Koyama and Goto [29] implemented a character animation optimizer controlled by editable parameters from users, and Sok et al. [61] presented a trajectory optimization to apply the changes of momentum and force to dynamic human motion. Carvalho et al. [13] proposed an optimization function for the latent motion space to edit the full-body motion by considering multiple key frame and trajectory constraints. Moreover, researchers have employed deep learning techniques in animation synthesis. Aberman et al. [1] introduced a method of motion style transfer based on a trained network without the need for paired training data. The proposed style transfer can adopt the motion style from the input motion by using the directly extracted motion style from the imported video. Li et al. [32] presented a generative model for motions that do not require pre-training but still provide high-quality and high-fidelity synthesized motions. Tang et al. [62] introduced an online framework that facilitates real-time motion generation supported by the trained network model.

Some researchers have focused on synthesizing and editing hand animation or facial animation. Ye and Liu [69] presented a randomized sampling algorithm synthesizing a set of hand motions manipulating the given object. Researchers [23, 41–43] proposed a method to synthesize hand motions based on the given body motions through the motion database composed of pairs of pre-captured finger and body motions. In the case of facial animation, Berson et al. [10] applied the convolutional neural network to overcome the issues of consistency between facial animations and highly frequent facial movement. Reed and Cosker [55] integrated the concept of evolutionary algorithms into facial animation synthesis. Their system sampled numerous facial expressions and asked users to select the appropriate facial expressions from the sampling. Through iterations, the system could synthesize facial animations as the evolved result based on users' guidance.

2.2 Character animation editing tool

Building upon the foundation provided by character animation synthesis and editing techniques, researchers have conducted studies to enhance user interactions and experience with character animation editing tools. Mukai and Kuriyama [45] introduced an editing system with a motion sequence visualization on a timeline and integrated drag-and-drop operation into the system to provide intuitive user interactions. Researchers have explored character animation editing based on various devices. Ciccone et al. [14] integrated several capture devices, such as hand-tracking sensors or full-body motion capture suits, to create the motion cycle intuitively. Cui and Mousas [18] introduced a system to control the virtual character's motion in real-time based on the input from motion capture sensors, such as the Leap Motion or Kinect. Also, Rhodin et al. [56] mapped wave properties of signals from body and hand trackers to non-human character skeletons to control their animations in real-time.

2.3 Animated pedagogical agents

Researchers have investigated animated pedagogical agents' effectiveness in education [22, 36, 60]. Annetta and Holmes [4] found that animated pedagogical agents can improve students' attitudes toward online lectures. Moreover, Poggiali [54] reported that animated videos with agents helped student hold their attention. The features of animated pedagogical agents impacting educational values have also been explored extensively [37], including visual presence [57], non-verbal communication [9], and communication style [65]. Mayer and Dapra [38] found that fully embodied agents with human-like gestures and appealing voices provided better educational outcomes for students than other agents communicating without human-like gestures. Also, Cui et al. [17] indicated that multimodal agents drove better learning than agents with only one channel. These features of the animated pedagogical agents can facilitate social connectivity and influence the students' learning environment. Gulz and Haake [19] reported that female students preferred the agent developing social connections to other agents providing only learning content.

2.4 Contribution

We developed an animation system that transforms online lectures into 3D animations using a pedagogical agent. This system provides animation editing tools to make the pedagogical agent more engaging. We evaluated our system with non-expert and expert groups, confirming our system's functionality, usability, and potential educational value. Our system can assist instructors and students who wish to enhance lecture videos for improved educational outcomes, including better lecture delivery, without requiring prior knowledge of animation or video editing techniques. Furthermore, our results offer valuable insights for researchers and developers interested in exploring user interaction and user experience in applications related to educational content.

3 Implementation details

This section presents the details of the implementation of our system. We developed our animation system as an extension of the Unity games engine (version 2021.3.3f1) with the Unity Registry packages (Barracuda, Timeline, and Unity Recorder) and used a Dell Alienware Aurora R7 desktop computer (Intel Core i7, NVIDIA GeForce RTX 2080, 32GB RAM). We discuss the implementation and functionalities in the following subsections.

3.1 Data extraction and animation enhancement

We implemented a system using third-party dependencies: `ThreeDPoseUnityBarracuda`[1] from Digital-Standard,

[1] https://github.com/digital-standard/ThreeDPoseUnityBarracuda.

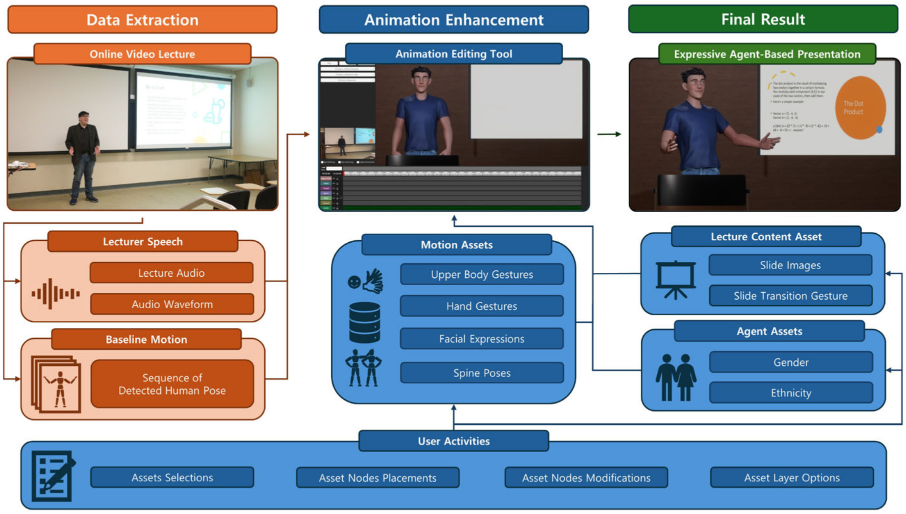


Fig. 1 The overview of our system. It consists of two main steps: data extraction and animation enhancement. After the animation enhancement step, the user can export the agent-based presentation in video format

SALSA LipSync Suite,[2] and Unity Timeline.[3] We divided our system into two steps: data extraction and animation enhancement. These two independent steps allow users to convert an existing video lecture into a 3D animation and then enhance this animation in terms of educational outcomes. We illustrate the two steps of our system in Fig. 1.

In the data extraction step, our system extracts the lecturer's motion from the instructional video as the baseline motion and converts it into a format compatible with our system. To do so, we integrated `ThreeDPoseUnity Barracuda`, which is an open source library with a trained model for human pose estimation. It supports the Kalman filter [66] with parameters for the covariance of the process noise (Q_k) and observation noise (R_k) to filter noises from the estimated human poses. Note that we set these parameters at $Q_k = .000125$ and $R_k = .0015$ in all examples we present in the supplementary video. When our system begins, the trained model estimates a human pose from each video frame. After the trained model completes human pose estimation for the whole video, our system generates a sequence of estimated human poses and converts it into an animation clip for the baseline motion. Also, our system analyzes the lecture's audio and extracts sample data to draw a waveform. The waveform presents the amplitude of the audio and provides a visual indication to the user about the audio accompanying the motion data.

In the animation enhancement step, our system retargets the extracted baseline motion to a pedagogical agent in a virtual classroom model, including a podium with a virtual computer and a slide frame. Then, using our custom-designed GUI-based animation editing tool, our system allows users to enhance the baseline extracted motion by applying additional short motion clips from a provided dataset, such as body gestures or facial expressions. Additionally, the users can adjust the properties of applied motion clips, such as position, length, and weight. Our system allows adding slides from the imported online video lecture with a pre-defined motion to support lecture delivery. Furthermore, the users can activate or deactivate eye movements. The users can change the pedagogical agent to reflect diversity based on different ethnicities and genders. Lastly, our system supports a function to play and pause the intermediate result. This function helps the users compare it with the original instructional video. After the users complete the animation enhancement, they can save their work and export an engaging agent-based presentation video.

3.2 Asset database

We developed an asset dataset to provide users with several resources to enhance the extracted motion sequence and educational outcomes. The asset database consists of motion, lecture content, and agent assets.

The motion assets consist of animation clips designed to make the pedagogical agent more engaging and improve its lecture delivery. These motion assets include four types of predefined motions: upper body gestures, hand gestures, facial expressions, and spine poses. We further divided the upper body gestures into affective, beat, and deictic gestures. For affective gestures, we extracted motions from the Geneva Multimodal Expression Corpus for Experimental Research on Emotion Perception (GEMEP) Core Set (Full Body) [6] using the same process as our pipeline [46]. These gestures from the GEMEP dataset convey six emotions: hot anger, disgust, fear, interest, elated joy, and sadness. Our animator identified five recurring beat gestures from Ted Talks for

[2] https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442.

[3] https://docs.unity3d.com/Packages/com.unity.timeline@1.2/manual/index.html.

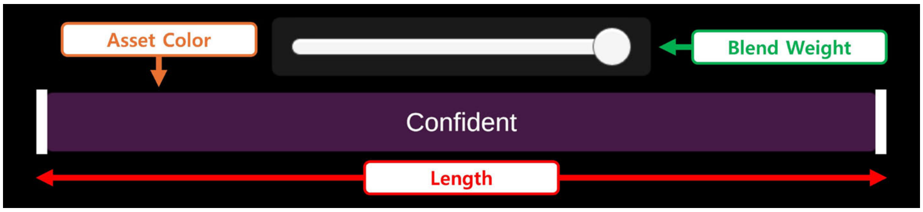


Fig. 2 An example of an asset node. We assigned a unique color to each type of asset node and allowed users to modify the blend weight and length by GUI-based interactions

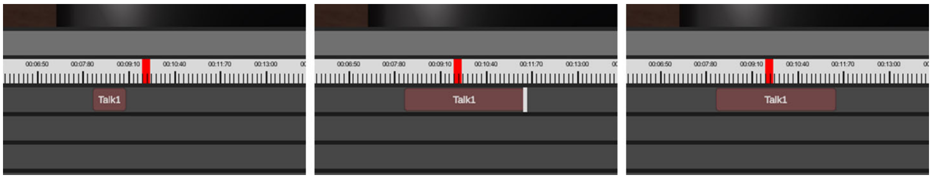


Fig. 3 Our system allows users to adjust the duration of an asset between short (left) and long (right) by dragging its ends (middle)

deictic gestures and recreated them using Reallusion's Character Creator 4 software. We synthesized hand gestures by referencing hand poses from the identified beat gestures.

We also animated nine facial expressions ranging from positive-active to positive inactive and from negative-active to negative-inactive emotions, such as confidence (positive-active). As for spine poses, we created six variations: back, forward, left-tilted, left-twisted, right-tilted, and right-twisted spine poses. These motion assets aimed to improve the naturalness and emotional expressiveness of the agent and provide variations by overriding the baseline motion or other motions from the asset database.

Regarding lecture content assets, they included lecture slide images and a fixed motion that simulates the action of pressing a button to change a lecture slide. Users can import their slide images for the lecture slides. Fixed motion was predefined by simulating the virtual character pressing a button on the virtual computer through inverse kinematics.

Lastly, the agent assets consist of 3D models of pedagogical agents representing ten different gender and ethnicity combinations to support their diversity and user preferences [8]. These models encompass two genders, male and female, and five ethnicities: Asian, Black, Hispanic, Indian, and White. In a previous study, Zhao et al. [71] validated the integrated pedagogical agent.

3.3 Asset Node

The asset node represents a specific asset applied to the baseline motion or the slide in the virtual environment during the animation enhancement process. Each asset node possesses three properties: length, weight, and color (see Fig. 2). The length of the asset node indicates the duration for which someone applied the asset. In the case of weight property (blend weight), it applies only to assets with motion and determines the extent to which motion overrides other motions. Lastly, the color of the asset node denoted its type. For example, we assigned purple to the facial expression asset node.

Users can modify the length and weight of the asset node. To adjust the length, users can drag each end of the asset node to increase or decrease its duration. For example, moving the right tip of the asset node to the right side will extend the duration (see Fig. 3), while moving it to the left side will shorten it.

Additionally, users can right-click on the asset node to activate the blend weight slider, enabling them to adjust the blend weight value. When users modify the blend weight, the system regenerates the motion curve associated with the asset node, ensuring that its maximum value corresponds to the blend weight value. The system blends the baseline motion with the motion of one asset node based on the value from its curve. For instance, if the motion from the asset node is an open arms gesture, and users move the slider to the middle, the system generates a motion curve with a maximum value of .50, causing the pedagogical agent to perform the half-open arms gesture (see Fig. 4).

3.4 Asset layers

We designed asset layers to visualize the animation enhancement progress and properties of assets. Our user interface offers two types of asset layers: the motion and the resource layers. The motion layer includes layers for upper body gestures, hand gestures, facial expressions, and spine poses. The resource layer encompassed layers for slide transitions, lip-sync animations, and audio (see Fig. 5).

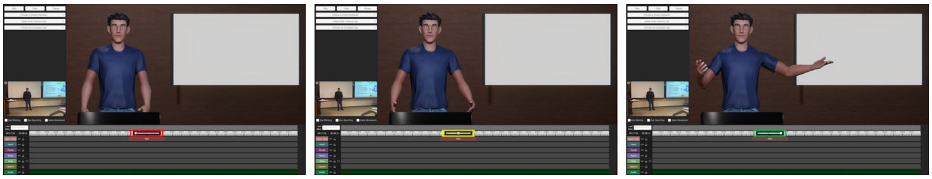


Fig. 4 Our system allows users to adjust the weight to blend the motion of the asset node with the baseline motion: 0% blend weight (left), 50% blend weight (middle), and 100% blend weight (right)

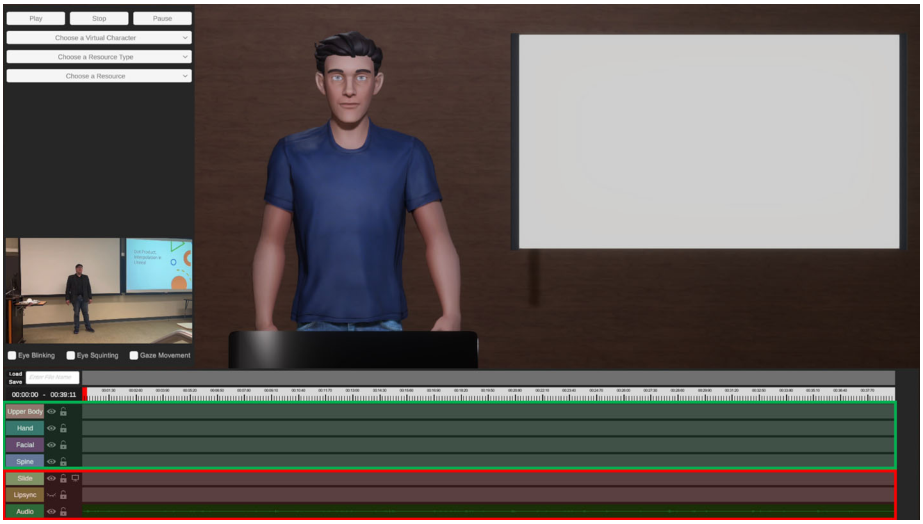


Fig. 5 Our system provides two types of layers: the motion (see highlighted area in green) and the resource (see highlighted area in red) layers

3.4.1 Motion layer

The motion layer enables users to blend the baseline motion with the motions of asset nodes by dragging and dropping asset nodes from the asset selector onto the corresponding motion layer. Initially, each motion layer has the same default color, but when users drag an asset node, the system changes the target motion layer's color to match the asset node's assigned color. This color change indicates which layer users should drop the asset node in. Additionally, each motion layer offers two options: a motion activation option that enables users to deactivate or activate the layer, controlling whether the layer applies to the baseline motion for easy comparison with user modifications, and a lock option that enables users to lock the layer to prevent errors during the animation enhancement step (see Fig. 6).

3.4.2 Resource layer

Each resource layer offers different functionality based on its properties. In the slide transition layer, users can add slides and apply a predefined motion for slide transitions to the pedagogical agent. By dragging the endpoints of the corresponding asset node, they can also control the duration for which each slide displays. Unlike the motion layers, it does not affect the length of the motion of the pedagogical agent for slide transitions. The slide transition layer provides activation and lock options similar to the motion layer. It also includes the slide activation option to control whether the slide should be displayed (see Fig. 7).

Both the lip-sync animation layer and the audio layer only visualize their properties. The lip-sync animation layer offers two different types: predefined lip-sync animations and real-time synthesized lip-sync animations using the SALSA LipSync Suite. It features a single option to toggle between these types of lip-sync animations. On the other hand, the audio layer displays the audio waveform based on the audio analysis from the data extraction step to assist users in identifying when the lecturer emphasizes specific points. It provides one option to mute or unmute the audio from the lecture.

3.5 Auxiliary user interface

Our system provides several user interface components to support the animation enhancement step. The timeline displays the duration of the baseline motion and includes an indicator to show users which part of the motion they enhance. Users can navigate to specific timesteps they want to check by dragging the indicator or pressing the play, stop, and pause buttons. Additionally, our system supports a save and load function, allowing users to preserve their work and reuse it later. Finally, our system offers three options to turn eye movements on or off: blinking, squinting, and gaze movement.

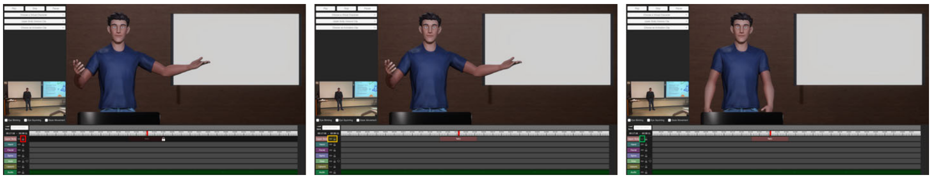


Fig. 6 Our system provides two options for the motion layer (middle; see highlighted area in yellow): lock option (left; see highlighted area in red) and activation option (right; see highlighted area in green). The left image shows the visual effect on the locked motion layer, and the right image shows the pedagogical agent with the baseline motion due to the inactivated motion layer

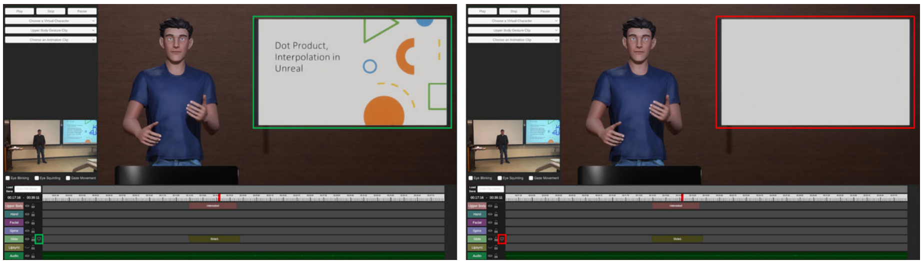


Fig. 7 The slide layer has the same options as motion layers and an extra option: slide activation. Users can make slides visible (left; see highlighted areas in green) or invisible (right; see highlighted areas in red)

4 User study

In this section, we provide the details of the user study we conducted.

4.1 Participants

We recruited two different groups of participants to evaluate our animation system: non-experts and experts. We recruited eight participants (age: $M = 28.25$, $SD = 3.45$) in the non-expert group. Of the non-expert sample, five were male (age: $M = 29.00$, $SD = 4.30$), and three were female (age: $M = 27.00$, $SD = 1.00$). They comprised graduate students with some prior teaching experience ($M = 1.32$, $SD = 1.18$). This group did not have expertise in animation; however, the participants had previous experience in video editing ($M = 3.38$, $SD = 2.67$). We recruited eight participants (age: $M = 47.43$, $SD = 8.30$) in the expert group, including people with teaching experience and computer graphics knowledge. Of the expert sample, seven were male (age: $M = 48.50$, $SD = 8.55$), and one was female (age: 41.00, $SD = .00$). They comprised professors from the department specializing in computer graphics and had considerable experience in animation ($M = 10.44$, $SD = 10.27$) and video editing ($M = 12.25$, $SD = 7.87$).

4.2 Experimental setup and measurements

We set up the experiment environment to support our participants in focusing on their user experience without any extraneous variables. We provided a QHD 27-inch monitor, Dell Alienware Aurora R7 desktop computer (Inter Core i7, NVIDIA Geforce RTX 2080, 32GB RAM), and stereo speakers in front of the participants. In the main study, we provided the pedagogical agent mimicking the lecturer from an 80-second online lecture video with duration.

To evaluate the usability of our system and how our participants perceived workload while using it, we used two questionnaires: the System Usability Scale (SUS) [11] and NASA Task Load Index (NASA TLX) [20]. For the SUS questions, our participants responded on a 5-point scale, and for the NASA TLX, they responded on a 7-point Likert scale. These questionnaires have been mainly employed in previous studies for system usability and task load evaluation [16, 67, 68]. Also, we conducted semi-structured interviews to explore their user experience more deeply. We developed our open-ended questions inspired by previously published work [3, 39, 49, 63]. The interview comprised ten questions about initial impression, ease of use, functionality, learning curve, comparison with other tools, the effectiveness of pedagogical agents, flexibility, technical issues, overall educational value, and suggestions for improvement. We provide the interview questions in Table 2 in Appendix A.

4.3 Procedure

When the participants arrived at the experimental site, we provided the consent form with the necessary information, such as the experimental procedure. We asked our participants to sign the form if they agreed with it. Note that our university's Institutional Review Board (IRB) approved our study. Then, the participants provided the demographic data

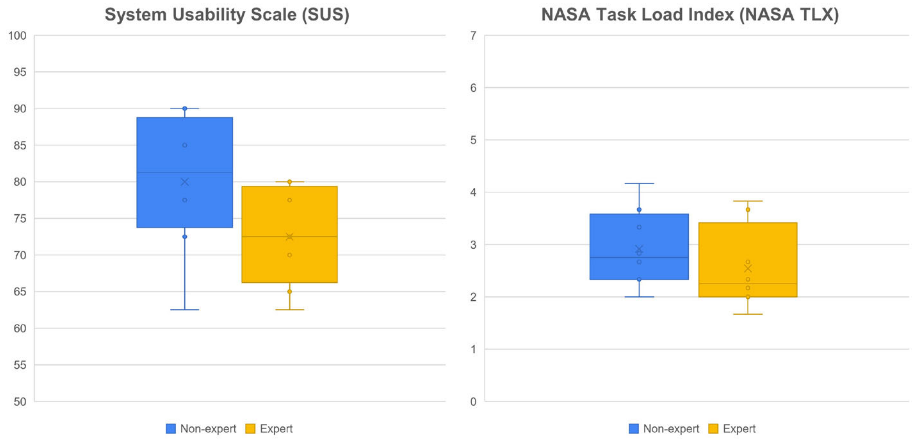


Fig. 8 Boxplots of quantitative data. There were no significant differences between the non-expert and expert groups. All SUS (left) scores are higher than the scale's mean. Regarding NASA TLX (right), the average scores of both groups are lower than the mean of the scale

and their years of experience in animation and video editing. Next, the researcher showed the tutorial video to the participants. The tutorial video provided instructions on how to use our system's functionalities. Our participants could watch the video as many times as needed and ask the researcher questions until they fully understood. Once our participants were ready to start the main study, the researcher asked them to edit the pedagogical agent to enhance its lecture delivery and educational outcomes through our system. The minimum required duration for the main study was 10 min, and most participants engaged with the system for more than 15 min. After the main study, the researcher asked the participants to answer the SUS and NASA TLX questionnaires. After the participants had responded to the questionnaire, the researcher interviewed them using predefined questions and did not limit their answers or interview time. Note that the interview took no more than 20 min. The researcher recorded the interview section to analyze the user experiences of participants. We also note that we did not compensate our participants and that they did not spend more than 60 min to complete the study.

5 Result

This section presents the details of the data analysis and its results.

5.1 Data analysis

We followed Brooke et al.'s instructions [11] to calculate the SUS score. For the NASA TLX, we calculated its mean of scores without weights. In the case of the interview, we converted the recorded interview audio file to the script through Naver CLOVA Note.[4] We analyzed the scripts based on the qualitative coding approach by Tastan et al. [63]. The approach comprises three steps: open, axial, and theoretical coding. In the open coding, we read the scripts and assigned a priori codes as annotations to the reportable responses. We repeated the open coding step multiple times to explore coded concepts and the common a priori codes between our participants. After completing the first step, we did the axial coding to categorize the coded concepts into themes. Then, we moved to theoretical coding to refine the themes as criteria for evaluating our system. Consequently, we established three concepts: educational value, functionality, and usability.

5.2 Quantitative data

To compare the two groups (see boxplots in Fig. 8), we used independent-sample t-tests. We did not find a statistically significant difference in the SUS score between the non-expert ($M = 80.00$, $SD = 9.45$) and expert ($M = 72.50$, $SD = 6.55$) groups; $t(14) = -1.845$, $p = .086$. For the NASA TLX, our statistical analysis did not reveal any significant difference between the non-expert ($M = 2.92$, $SD = .75$) and expert ($M = 2.54$, $SD = .81$) groups; $t(14) = -.970$, $p = .349$.

5.3 Qualitative data

To demonstrate reported strength, limitation, and improvement, we categorized the concepts and themes from qualitative coding to positive concepts, negative concepts, and improvement (see Table 1). Our participants, including non-experts (NEP) and experts (EP), mentioned 24 concepts 199 times. Specifically, positive concepts were the most frequently mentioned (133 times), followed by improvements (37 times). Negative concepts were the least frequently mentioned (29 times). These findings indicated that our system provided positive user experiences to our participants.

[4] https://clova.ai/speech/en.

Table 1 Reported positive and negative concepts and improvements to our system

Theme	Concepts	Overall		Non-expert		Expert	
		Freq	Nop	Freq	NoP	Freq	Nop
Positive concepts							
Educational value	CEP1: Potential to enhance education outcomes.	11	10	6	5	5	5
	CEP2: Improve the lecture delivery.	6	5	5	4	1	1
	CEP3: Paying more attention to the lecture.	9	7	6	4	3	3
Functionality	CFP1: Capability to synthesize customized gestures	15	12	7	6	8	6
	CFP2: Usefulness of pre-defined animation clips.	9	7	8	6	1	1
	CFP3: Variety of pre-defined animation clips	4	3	2	1	2	2
	CFP4: Efficient and effective editing functions	6	4	3	2	3	2
Usability	CUP1: Require minimal effort to learn	40	16	20	8	20	8
	CUP2: Clear layout and visual support	14	9	5	4	9	5
	CUP3: Intuitive GUI-based interactions	19	10	8	6	11	4
Negative concepts							
Educational value	CEN1: Prefer the human lecturer rather than the pedagogical agent	2	2	0	0	2	2
	CEN2: Necessary to support various preferences of the pedagogical agent	2	2	1	1	1	1
Functionality	CFN1: A limited number of pre-defined animation clips	5	3	5	3	0	0
	CFN2: Fixed perspective of the 3D animation	3	3	0	0	3	3
	CFN3: Unable to edit motions while the 3D animation is playing	5	5	4	4	1	1
Usability	CUN1: Require repeated manual editing	6	5	4	3	2	2
	CUN2: Difficulties in navigating the timeline	6	5	3	3	3	2
Improvements							
Educational value	CEI1: Provide more pedagogical agents	3	3	1	1	2	2
Functionality	CFI1: Provide more pre-defined animation clips	13	9	5	4	8	5
	CFI2: Support different camera perspectives	4	3	0	0	4	3
	CFI3: Provide lecture transcripts.	2	2	0	0	2	2
	CFI4: Support overridden asset layers.	3	2	3	2	0	0
Usability	CUI1: Automatic asset node placement and manipulation.	7	3	0	0	7	3
	CUI2: Provide functions to navigate the timeline.	5	4	2	2	3	2

"Freq" denotes how many responses our participants provided for the corresponding concept, and "NoP" denotes the number of participants who mentioned the corresponding concept. For the terms CEP, CEN, or CEI, the first letter (C) denotes that it is a concept, the second letter (E, F, and U) indicates its theme (Educational value, Functionality, and Usability), and the last letter (P, N, and I) denotes the type of concept (Positive, Negative, and Improvement)

5.3.1 Positive concepts

Our participants mentioned positive concepts about the educational value of our system (CEP1-CEP3) 26 times. Ten participants addressed the potential of our system to enhance education outcomes (CEP1) 11 times. This finding indicates that our participants agreed that the 3D animations synthesized using our system could provide educational benefits. Specifically, some participants focused on the context of the distance learning environment. For example, NEP8 reported: *"...I think this could be a useful tool for education, especially for distance learning."* Also, five participants mentioned that our system improved the lecture delivery (CEP2) six times. NEP5 said: *"It (the pedagogical agent) can help in virtual classes a lot. I am taking my stat classes virtually right now, and I think my stat professor does not have good body language (unlike the pedagogical agent)."* Last, nine participants said that the output of our system could help students pay attention to the lecture (CEP3) nine times. For instance, EP8 said: *"...I felt like the agent really focused me on the content so pedagogically my focus was well on the dot product lesson..."*

Regarding functionality, our participants reported positive concepts (CFP1-CFP4) 34 times. Twelve participants mentioned our system is excellent in synthesizing upper body gestures (CFP1) 15 times. It shows that our system provided flexibility and customizability in catering to different educational content. For instance, EP1 said: *"...so I think it already provides a lot of options to create this custom animation,"* and NEP1 reported: *"...because like the added hand gestures and body gestures make it easy to address animations for different topics..."* Also, seven participants reported the usefulness of predefined animation clips (CFP2) nine times. In particular, they highlighted the importance of animation clips for the upper body and facial expression. NEP1 mentioned: *"The most important features actually highlighted the upper body and facial expressions."* Furthermore, three participants mentioned the variety of predefined animation clips (CFP3) four times. EP7 said: *"I like the separation between the different types of body animation..."* Last, four participants reported the efficient and effective editing functions (CFP4) six times. They highlighted the blending function between animations with the weight. EP6 said: *"I can make some of the blending between the two animations, and then that would be the most helpful way to generate the natural animation."*

Last, our participants mentioned positive usability concepts (CUP1-CUP3) 73 times. Sixteen participants reported that our system requires minimal effort to learn (CUP1) 40 times, which means all participants agreed that our system is easy to learn. Some participants compared our system with other applications, and others focused on the non-expert's ease of use. For example, NEP7 said: *"...it is definitely much easier to use compared to Maya,"* and EP8 mentioned: *"I think even for a non-expert, I think it would be. I think most people would understand this."* Also, nine participants mentioned the clear layout and visual support (CUP2) 14 times. Some participants reported the similarity with other tools, and others focused on the asset node's assigned color and highlighted the asset layer during the interaction. EP1 mentioned: *"...very similar to the workspace I used for the other tools,"* and EP7 mentioned: *"...everything here is color-coded."* Last, 10 participants reported the intuitive GUI-based interaction (CUP3) 19 times and usually mentioned dragging and dropping asset nodes to the asset layer as an example. NEP7 said: *"...this drag and drop workflows are super user friendly,"* and EP8 mentioned: *"...the timeline and the ability to drop directly that seems very natural.*

5.3.2 Negative concepts

Although our system received a positive evaluation, there were also negative aspects. Our participants reported negative concepts (CEN1 and CEN2) five times regarding education value. Two participants pointed out that they preferred the human lecturer over the pedagogical agent (CEN1) two times. EP2 said: *"...I feel like more detached from the virtual (pedagogical) agent than I would feel from the professor even if the professor was boring. At least I know that they are real."* Also, two participants addressed the need to support various preferences of the pedagogical agent (CEN2) two times. For example, EP4 mentioned: *"...so I guess I feel like hyper-realistic would be better... it would probably be an individual preference."*

As for functionality, our participants reported its negative concepts (CFN1-CFN3) 13 times. Three participants mentioned the limited number of predefined animation clips (CFN1) five times. NEP2 said: *"The current number of animations we have, I would not say it is enough."* Furthermore, three participants addressed the fixed perspective of the 3D animation (CFN2) three times. Specifically, they pointed out when the students needed to focus on the slide image rather than the lecturer. For example, EP7 mentioned: *"...I think sometimes students might want to zoom in on the slide if the text is too small."* Last, five participants pointed out they could not edit motions while the 3D animation was playing (CFN3) five times. NEP7 said: *"I cannot edit the animation while I am playing (the animation)."*

Last, our participants mentioned negative concepts about usability (CUN1 and CUN2) 12 times. Five participants said our system required repeated manual editing (CUN1) six times. Most participants pointed out that finding assets through an asset selector was cumbersome. NEP5 said: *"Dropdown things (asset selector)... are not very efficient."* Also, six participants mentioned difficulties in navigating the timeline (CUN2) six times. For instance, NEP4 mentioned:

"...I need to manually drag it (timeline indicator), and I feel like, oh, that will spend a little more time."

5.3.3 Improvements

During the interview, our participants suggested improving our system to make it more effective and user-friendly. The improvements in educational value include one concept, CEI1. Three participants proposed to provide more pedagogical agents (CEI1) three times. EP3 said: *"...it is some type of animal or something that would be neat..."*

Regarding functionality, its improvements are composed of four concepts, from CFI1 to CFI4, and were reported 22 times. Nine participants suggested providing more predefined animation clips (CFI1) 13 times. Some participants mentioned the necessary animation clips, such as walking or eye expressions, to enhance the learning experience. For instance, NEP6 said: *"...the eye movement and all that if they can get more fluid then I think that would make the learning experience even better."* Also, three participants proposed to support different camera perspectives (CFI2) four times. Specifically, they mentioned that it would help students focus on the context if one of the camera perspectives provided only the slide image. For instance, EP2 said: *"...you can just have the character sometimes a kind of like a close-up, and then the rest of it is just slides. Not so much focus on the character but in the context."* Furthermore, two participants suggested providing lecture transcripts (CFI3) two times. They mentioned that the transcripts could reduce the time spent finding specific moments in the timeline for motion editing. For example, EP3 said: *"That (transcript) would be more useful to know the different points in the lecture when I might want to change an animation."* Lastly, two participants proposed supporting overridden layers for each motion layer (CFI4) three times. NEP2 mentioned: *"...if I want to blend the two animations, it would be really nice if we had that feature (overridden layers)."*

Last, the usability improvements include two concepts, CUI1 and CUI2, which our participants reported 12 times. Three participants suggested automatic asset node placement and manipulation (CUI1) seven times. EP6 said: *"...you should not really need to have the task of dragging it (asset node) into the timeline. If you click, it should just go to the next thing in the timeline."* Also, four participants proposed that functions to navigate the timeline (CUI2) should be provided five times. The proposed functions are composed of zoom in and out, thumbnails of the timeline indicator, annotations on the timeline, and click-based timeline navigation. EP5 mentioned: *"...sometimes the timing is hard. So, I think you should have a system (timeline) for marker."*

5.3.4 Different responses between groups

The non-expert group provided the majority of responses from CEP2. For example, NEP3 mentioned: *"I think it is going to enhance learning because each user can now customize how they want the instructor in their videos to demonstrate to them."* However, unlike the non-expert group, the expert group wondered whether the pedagogical agent could provide more educational benefits than human lecturers in online video lectures. In the case of CEN1, all responses were from the expert group. EP4 said: *"I would have to see if there was some pedagogical benefit to using the animated character over using the video from an educational standpoint."*

Regarding functionality, the non-expert group provided all responses from CFN1 and most responses from CFP2 and CFI4. For instance, NEP5 said: *"I think the most important feature was the upper body one (motion),"* and NEP6 mentioned: *"...definitely the facial expressions and hand gestures were stuff I felt was important."* In contrast, the expert group provided most responses from CFN2, CFI2, and CFI3. For example, EP6 said: *"The slides I think the slides are really important, and they are relegated to a very small section of the screen compared to the character."*

Regarding usability, the expert group provided all responses from CUI1, which showed that the expert group focused on solving repeated manual editing (CUN1). Additionally, the frequency of improvements from the expert group was more than twice the frequency of improvement from the non-expert group. For instance, EP1 mentioned: *"...it is kind of like a Chinese version of YouTube. And when they upload the videos, the system can automatically detect different chapters inside the video."*

6 Discussion

Based on our results, we found no significant differences from the quantitative data analysis between the two groups. However, we found reportable findings by interpreting the SUS scores based on the adjective rating by Bangor et al. [5]. Specifically, we converted these scores to the predefined adjectives: "worst imaginable," "awful," "poor," "ok," "good," "excellent," and "best imaginable." Not only did the overall group surpass the score of "good" ($M = 71.4$), but each group also achieved scores above this score. These findings indicated that our system is usable regardless of users' knowledge and experiences. Similarly, in the case of NASA TLX, we found the overall scores and scores from each group were lower than the mean of the scale ($M = 3.50$). Hence, these findings showed our participants perceived a low task load during our system experience without regard to their experience and knowledge.

From the qualitative analysis, we explored our system's positive and negative concepts and improvements in educational value, functionality, and usability. On the one hand, positive concepts were the most frequent; on the other hand, negative concepts were the least frequent. This finding indicated that our system helped users improve the lecture's educational outcomes and provided enough capabilities to customize the lecture and intuitive user experiences. Specifically, all participants mentioned that our system requires minimal effort to learn. This finding stated that our system has a user-friendly user interface and interaction design regardless of users' experience or knowledge of computer graphics.

Also, we found some biased responses from the comparison between groups. For educational value, the non-expert group provided most responses from CEP2, and the expert group provided all responses from CEN1. These findings indicated that the non-expert group thought the pedagogical agent could deliver the lecture better; on the contrary, the expert group wondered about the effectiveness of the pedagogical agent when comparing it with human lecturers.

In the case of functionality, the non-expert group provided all responses of CFN1 and most responses of CFP2 and CFI4. These findings indicated that the non-expert group focused on the motion of the pedagogical agent rather than other features, such as camera perspective or lecture contents. In contrast, the expert group provided the most responses from CFN2, CFI2, and CFI3. These findings showed that the expert group prioritized the lecture content and format of 3D animation rather than the motion of the pedagogical agent, and this finding aligns with the findings from the educational value.

Last, we also found some biased responses in usability. The expert group provided all responses from CUI1, showing that the expert group focused on solving repeated manual editing (CUN1). Additionally, the frequency of improvements from the expert group was more than twice the frequency of improvement from the non-expert group. This finding indicated that the expert group provided more comments about the improvement based on their knowledge and experiences.

7 Limitation

When evaluating a system, it is necessary to check its capabilities and study design. Although our participants engaged our system without encountering any critical issues, we would like to report the limitations of our study. Note that these limitations do not invalidate our system and the reported results. Instead, they inform researchers who are interested in developing similar animation software.

First, the quality of baseline motion was not promising due to the limited capability of the integrated third-party model. The `ThreeDPoseUnityBarracuda` could extract the lecturer's pose data from the imported video, but the extracted pose data had too much noise. Although it provided the Kalman filter for denoising, the denoised motion looked too smooth and a bit far from the motion of the lecturer. Also, the lip-sync animation from SALSA LipSync Suite had delays from the lecture video, and its quality is not enough when compared with lip-sync animation synthesized by experienced animators. This could have negatively impacted our participants' evaluation of our system.

Second, our participants were exposed to only a male lecturer and pedagogical agent because we used one video for the experiment, and it was necessary to match the gender of the pedagogical agent and the lecturer's voice. Unfortunately, according to Makransky et al. [35], gender can be one of the factors affecting learning outcomes. So, in terms of the educational outcomes, there is a chance for the evaluation of our system from male and female participants to be slightly different.

Third, our participants mainly reported the limited number of predefined animation clips as improvements to our system. It is apparent that more animation clips help users synthesize more gestures and lead to educational benefits. So, although we provided numerous animation clips to customize pedagogical agents' gestures, they thought it would be better if they had more.

Fourth, we provided a limited number of animation controls. Although our participants evaluated the functionality of our system as efficient and effective, it would have been beneficial if we had provided other animation controls, such as replacing an animation clip with another clip or highlighting a specific spot on the motion and resource layers to indicate where the lecturer emphasized based on audio analysis. Also, we did not provide any function to edit rigs of pedagogical agents or synthesize animations by adjusting joint or inverse kinematics. We argue that synthesizing customized animations might be limited compared to other animation editing tools, such as Autodesk's Maya.

Last, our sample size from the experiment was not enough. Although we employed both quantitative and qualitative research methodologies, the result from the quantitative analysis would be more convincing if we had more participants.

8 Conclusion and future work

We presented an animation system that converts instructional videos to pedagogical agent-based presentations and provides animation editing tools to improve the expressiveness of the pedagogical agents. We evaluated our system's potential educational value, functionality, and usability through a user study with the non-expert and expert groups. We used quantitative and qualitative research methodologies and reported findings from questionnaires and interviews. The results of SUS and NASA TLX showed our system's good usability and low task load regardless of users' experience and knowledge in computer graphics and animation. Also, the findings from the interview highlighted our system's intuitiveness and capabilities to customize agent-based presentations and enhance educational benefits. Furthermore, we found interesting results when we compared the two groups. The non-expert group focused on the motions of the pedagogical agent; on the other hand, the expert group paid attention to the lecture contents rather than a pedagogical agent.

The study had some limitations, such as the low quality of baseline motion, biased gender of the pedagogical agent, and a limited number of animation clips. Therefore, as future works, we plan to add more animation clips and integrate other state-of-the-art third-party models into our system to improve the quality of extracting the baseline motion from videos. Also, we will implement a function that converts the lecturer's voice to that of the different genders to enable a variety of pedagogical agents regardless of the gender of the lecturer in the online video lecture. Moreover, we will implement more functionalities, such as a recommendation tool that automatically enhances the baseline animation with recommended motion clips from our database.

A Appendix

We developed interview questions to evaluate our animation system in terms of user experience, educational benefits, and functionality. We provide our interview questionnaire in Table 2.

Table 2 Interview questions we used in our study

Intent of questions	Interview questions
Initial impressions	What were your first impressions upon using our animation software, particularly in terms of its layout and overall design?
Ease of Use	How intuitive did you find the software when creating an animation for the first time? Were there any features or tools that were particularly easy or difficult to use?
Functionality	Which features of the software did you find most useful for creating educational animations, and why?
Learning curve	How steep was the learning curve for you in understanding how to use all the functionalities of the software? What aspects, if any, required more time to master?
Comparison with other tools	How does our software compare with other animation or video editing tools you've used, especially in the context of creating educational content?
Effectiveness of Pedagogical Agents	In your opinion, how effectively does the software integrate pedagogical agents into the animated lectures? Do you think these agents add value to the educational content?
Customization and flexibility	Did you find the software flexible and customizable enough to cater to different styles of educational content? Can you provide an example?
Technical Issues	Did you encounter any technical issues or limitations while using the software? How did these affect your user experience?
Overall educational value	Based on your experience, how would you rate the software's potential to enhance learning and engagement in an educational setting?
Suggestions for improvement	What improvements or additional features would you suggest to make this software more effective and user-friendly for educators and instructional designers?

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00371-024-03622-w.

Acknowledgements This project was supported by an NSF RETTL Award (Award Number: 2201019).

Author Contributions M.C. implemented the presented animation system. M.C., C.M., and N.A. wrote the main manuscript text. S.P. and K.H. synthesized animation clips and wrote the part about it from the manuscript text. F.Z. and R.M. wrote the part related to education in the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Aberman, K., Weng, Y., Lischinski, D., Cohen-Or, D., Chen, B.: Unpaired motion style transfer from video to animation. ACM Trans. Graph. (TOG) **39**(4), 64 (2020)
2. Alexanderson, S., Nagy, R., Beskow, J., Henter, G.E.: Listen, denoise, action! audio-driven motion synthesis with diffusion models. ACM Trans. Graph. (TOG) **42**(4), 1–20 (2023)
3. Ali, L., Hatala, M., Gašević, D., Jovanović, J.: A qualitative evaluation of evolution of a learning analytics tool. Comput. Educ. **58**(1), 470–489 (2012)
4. Annetta, L.A., Holmes, S.: Creating presence and community in a synchronous virtual learning environment using avatars. Int. J. Inst. Technol. Dist. Learn. **3**(8), 27–43 (2006)
5. Bangor, Aaron, Kortum, Philip, Miller, James: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)
6. Bänziger, T., Mortillaro, M., Scherer, K.R.: Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. Emotion **12**(5), 1161 (2012)
7. Basten, Ben, Egges, Arjan: Motion transplantation techniques: a survey. IEEE Comput. Graph. Appl. **32**(3), 16–23 (2011)
8. Baylor, A., Shen, E., Huang, X.: Which pedagogical agent do learners choose? The effects of gender and ethnicity. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 1507–1510. Association for the Advancement of Computing in Education (AACE), (2003)
9. Baylor, A.L., Kim, S.: Designing nonverbal communication for pedagogical agents: when less is more. Comput. Hum. Behav. **25**(2), 450–457 (2009)
10. Berson, E., Soladié, C., Barrielle, V., Stoiber, N.: A robust interactive facial animation editing system. In: *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pp. 1–10, (2019)
11. Brooke, John, et al.: SUS-a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
12. Cardle, M., Barthe, L., Brooks, S., Robinson, P.: Music-driven motion editing: Local motion transformations guided by music analysis. In: *Proceedings 20th Eurographics UK Conference*, pp. 38–44. IEEE, (2002)
13. Carvalho, S.R., Boulic, R., Vidal, C.A., Thalmann, D.: Latent motion spaces for full-body motion editing. Vis. Comput. **29**, 171–188 (2013)
14. Ciccone, L., Guay, M., Nitti, M., Sumner, R.W.: Authoring motion cycles. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 1–9, (2017)
15. Cook, David A.: The value of online learning and MRI: finding a niche for expensive technologies. Med. Teach. **36**(11), 965–972 (2014)
16. Cui, D., Mousas, C.: Exploring the effects of virtual hand appearance on midair typing efficiency. Comput. Anim. Virtual Worlds **34**(3–4), e2189 (2023)
17. Cui, J., Popescu, V., Adamo-Villani, N., Cook, S.W., Duggan, K.A., Friedman, Howard S.: Animation stimuli system for research on instructor gestures in education. IEEE Comput. Graph. Appl. **37**(4), 72–83 (2017)
18. Cui, Y., Mousas, C.: Master of puppets: an animation-by-demonstration computer puppetry authoring framework. 3D Res. **9**, 1–14 (2018)
19. Gulz, A., Haake, M.: Social and visual style in virtual pedagogical agents. In *Workshop: adapting the interaction style to affective factors, 10th International Conference on User Modelling (UM'05)*, (2005)
20. Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, (2006)
21. Horovitz, T., Mayer, R.E.: Learning with human and virtual instructors who display happy or bored emotions in video lectures. Comput. Hum. Behav. **119**, 106724 (2021)
22. Johnson, W.L., Lester, J.C.: Face-to-face interaction with pedagogical agents, twenty years later. Int. J. Artif. Intell. Educ. **26**, 25–36 (2016)
23. Jörg, S., Hodgins, J., Safonova, A.: Data-driven finger motion synthesis for gesturing characters. ACM Trans. Graph. (TOG) **31**(6), 1–7 (2012)
24. Jovane, A., Raimbaud, P., Zibrek, K., Pacchierotti, C., Christie, M., Hoyet, L., Olivier, A., Pettré, J.: Warping character animations using visual motion features. Comput. Graph. **110**, 38–48 (2023)
25. Kentnor, H.E.: Distance education and the evolution of online learning in the united states. Curric. Teach. Dialogue **17**(1), 21–34 (2015)
26. Kim, J., Kim, J., Choi, S.: Flame: Free-form language-based motion synthesis & editing. In: Proceedings of the AAAI Conference on Artificial Intelligence **37**, 8255–8263 (2023)
27. Kim, Jungjoo, Kwon, Yangyi, Cho, Daeyeon: Investigating factors that influence social presence and learning outcomes in distance higher education. Comput. Educ. **57**(2), 1512–1520 (2011)
28. Kovar, L., Gleicher, M.: Flexible automatic motion blending with registration curves. In: *Symposium on Computer Animation*, vol. 2. San Diego, CA, USA, (2003)
29. Koyama, Y., Goto, M.: Optimo: Optimization-guided motion editing for keyframe character animation. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, (2018)
30. Lawson, Alyssa P., Mayer, Richard E., Adamo-Villani, Nicoletta, Benes, Bedrich, Lei, Xingyuc, Cheng, Justin: Do learners recognize and relate to the emotions displayed by virtual instructors? Int. J. Artif. Intell. Educ. **31**, 134–153 (2021)
31. Lawson, A.P., Mayer, R.E., Adamo-Villani, N., Benes, B., Lei, X., Cheng, J.: Recognizing the emotional state of human and virtual instructors. Comput. Hum. Behav. **114**, 106554 (2021)
32. Li, Weiyu, Chen, Xuelin, Li, Peizhuo, Sorkine-Hornung, Olga, Chen, Baoquan: Example-based motion synthesis via generative motion matching. ACM Trans. Graph. (TOG) **42**(4), 1–12 (2023)
33. Loderer, K., Pekrun, R.: Emotional foundations of game-based learning. In: Handbook of Game-Based Learning, pp. 111–151. MIT Press, Cambridge (2020)

34. Lyu, Lei, Zhang, Jinling: Stylized human motion warping method based on identity-independent coordinates. Soft Comput. **24**(13), 9765–9775 (2020)
35. Makransky, G., Wismer, P., Mayer, R.E.: A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. J. Comput. Assist. Learn. **35**(3), 349–358 (2019)
36. Martha, A.S.D., Santoso, H.B.: The design and impact of the pedagogical agent: a systematic literature review. J. Educ. Online **16**(1), 1 (2019)
37. Mayer, R.E.: Multimedia Learning. Elsevier, Amsterdam (2020)
38. Mayer, R.E., DaPra, C.S.: An embodiment effect in computer-based learning with animated pedagogical agents. J. Exp. Psychol. Appl. **18**(3), 239 (2012)
39. Mills, R., Haga, S.B.: Qualitative user evaluation of a revised pharmacogenetic educational toolkit. Pharmacogen. Person. Med. **11**, 139–146 (2018)
40. Mousas, C., Anagnostopoulos, C.-N.: Chase: character animation scripting environment. In: VRCAI, pp. 55–62. Springer, Cham (2015)
41. Mousas, C., Anagnostopoulos, C.-N.: Learning motion features for example-based finger motion estimation for virtual characters. 3D Res. **8**, 1–12 (2017)
42. Mousas, C., Anagnostopoulos, C.-N.: Real-time performance-driven finger motion synthesis. Comput. Graph. **65**, 1–11 (2017)
43. Mousas, C., Anagnostopoulos, C-N., Newbury, P.: Finger motion estimation and synthesis for gesturing characters. In: *Proceedings of the 31st Spring Conference on Computer Graphics*, pp. 97–104, (2015)
44. Mukai, T., Kuriyama, S.: Geostatistical motion interpolation. In: *ACM SIGGRAPH 2005 Papers*, pp. 1062–1070. (2005)
45. Mukai, T., Kuriyama, S.: Pose-timeline for propagating motion edits. In: *Proceedings of the 2009 ACM siggraph/eurographics symposium on computer animation*, pp. 113–122, (2009)
46. Mukanova, M., Adamo, N., Mousas, C., Choi, M., Hauser, K., Mayer, R., Zhao, F.: Animated pedagogical agents performing affective gestures extracted from the gemep dataset: Can people recognize their emotions? In: *International Conference on ArtsIT, Interactivity and Game Creation*, pp. 271–280. Springer, (2023)
47. Neff, M., Kim, Y.: Interactive editing of motion style using drives and correlations. In: *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 103–112, (2009)
48. Nikopoulou-Smyrni, P., Nikopoulos, C.: Evaluating the impact of video-based versus traditional lectures on student learning. (2010)
49. Nikpeyma, N., Zolfaghari, M., Mohammadi, A.: Barriers and facilitators of using mobile devices as an educational tool by nursing students: a qualitative research. BMC Nurs. **20**, 1–11 (2021)
50. Oshita, M.: Smart motion synthesis. In: Computer Graphics Forum, vol. 27, pp. 1909–1918. Wiley, New York (2008)
51. Oshita, M.: Generating animation from natural language texts and semantic analysis for motion search and scheduling. Vis. Comput. **26**, 339–352 (2010)
52. Oshita, Masaki, Seki, Takeshi, Yamanaka, Reiko, Nakatsuka, Yukiko, Iwatsuki, Masami: Easy-to-use authoring system for Noh (Japanese traditional) dance animation and its evaluation. Vis. Comput. **29**, 1077–1091 (2013)
53. Pekrun, R., Stephens, E.J.: Achievement emotions: a control-value approach. Soc. Pers. Psychol. Compass **4**(4), 238–255 (2010)
54. Poggiali, J.: Student responses to an animated character in information literacy instruction. Lib. Hi Tech **36**(1), 29–42 (2017)
55. Reed, K., Cosker, D.: User-guided facial animation through an evolutionary interface. In: Computer Graphics Forum, vol. 38, pp. 165–176. Wiley, New York (2019)
56. Rhodin, H., Tompkin, J., Kim, K.I., De Aguiar, E., Pfister, H., Seidel, H.P., Theobalt, C.: Generalizing wave gestures from sparse examples for real-time character control. ACM Trans. Graph. (TOG) **34**(6), 1–12 (2015)
57. 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**(6), 2741–2756 (2008)
58. Rubenstein, H.: Recognizing e-learning's potential & pitfalls. Learn. Train. Innov. **4**(4), 38 (2003)
59. Sauer, D., Yang, Y.-H.: Music-driven character animation. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **5**(4), 1–16 (2009)
60. 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)
61. Sok, K.W., Yamane, K., Lee, J., Hodgins, J.: Editing dynamic human motions via momentum and force. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer animation*, pp. 11–20. Citeseer, (2010)
62. Tang, X., Wu, L., Wang, H., Hu, B., Gong, X., Liao, Y., Li, S., Kou, Q., Jin, X.: Rsmt: Real-time stylized motion transition for characters. In *ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1–10, (2023)
63. Tastan, H., Tuker, C., Tong, T.: Using handheld user interface and direct manipulation for architectural modeling in immersive virtual reality: an exploratory study. Comput. Appl. Eng. Educ. **30**(2), 415–434 (2022)
64. Wang, M., Chen, Z., Shi, Y., Wang, Z., Xiang, C.: Instructors' expressive nonverbal behavior hinders learning when learners' prior knowledge is low. Front. Psychol. **13**, 810451 (2022)
65. 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**(2), 98–112 (2008)
66. Welch, G., Bishop, G., et al.: An introduction to the Kalman filter. (1995)
67. Xu, X., Gong, J., Brum, C., Liang, L., Suh, B., Gupta, S.K., Agarwal, Y., Lindsey, L., Kang, R., Shahsavari, B., et al.: Enabling hand gesture customization on wrist-worn devices. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–19, (2022)
68. Xu, X., Yu, A., Jonker, T.R., Todi, K., Lu, F., Qian, X., Belo, J.M.E., Wang, T., Li, M., Mun, A., et al.: Xair: A framework of explainable AI in augmented reality. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–30, (2023)
69. Ye, Yuting, Liu, C.K.: Synthesis of detailed hand manipulations using contact sampling. ACM Trans. Graph. (ToG) **31**(4), 1–10 (2012)
70. Zhang, J.-Q., Xu, X., Shen, Z.-M., Huang, Z.-H., Zhao, Y., Cao, Y.-P., Wan, P., Wang, M.: Write-an-animation: high-level text-based animation editing with character-scene interaction. Comput. Graph. Forum **40**, 217–228 (2021)
71. Zhao, F., Mayer, R.E., Adamo-Villani, N., Mousas, C., Choi, M., Lam, L., Mukanova, M., Hauser, K.: Recognizing and relating to the race/ethnicity and gender of animated pedagogical agents. J. Educ. Comput. Res. **62**(3), 675–701 (2024)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

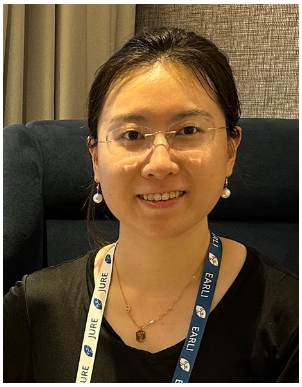
Minsoo Choi is a Ph.D. student in the Department of Computer Graphics Technology at Purdue University. His research interests include character animation, virtual reality, and HCI (human-computer interaction).

Christos Mousas is an Associate Professor in the Department of Computer Graphics Technology and Director of the Virtual Reality Lab at Polytechnic Institute, Purdue University. His research revolves around virtual reality, virtual humans, computer graphics & animation, intelligent systems, and human-computer interaction. He serves as an Associate Editor for the Computer Animation and Virtual Worlds and Frontiers in Virtual Reality journals as well as on the organizing and program committees of many conferences in the virtual reality, computer graphics/animation, and human-computer interaction fields.

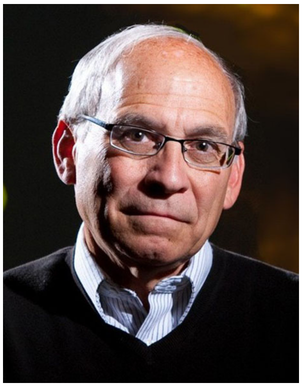
Nicoletta Adamo is a Professor of Computer Graphics Technology and Purdue University Faculty Scholar. She is an award-winning animator and graphic designer and creator of several 2D and 3D animations that aired on national television. Her area of expertise is in character animation and character design, and her research interests focus on the application of 3D animation technology to education, HCI (human-computer interaction), and visualization.

Sanjeevani Patankar is a Master's student in Computer Graphics Technology at Purdue University. Her research specializes in computer animation and character believability. After graduating from Ringling College of Art and Design, Sanjeevani aims to merge art and technology and push animation as a medium.

Klay Hauser is from West Lafayette Indiana. He studied at Purdue University for his Master's in computer graphics technology. He is currently a 1st-year PhD student at Purdue University. His research focuses on human perception of virtual agents, specifically in relation to facial expression asymmetries.



Fangzheng Zhao is a fifth-year PhD Candidate in Dr. Richard Mayer's Lab at the University of California, Santa Barbara. Her research focuses on multimedia learning, specifically exploring various strategies to improve the effectiveness of video or game-based learning.



Richard E. Mayer is Distinguished Professor of Psychology at the University of California, Santa Barbara. His research interests are in applying the science of learning to education, with current projects on multimedia learning, computer-supported learning, and computer games for learning.