

Effects of Superimposing Salient Graphics on Learning Material

Shu Matsuura^(✉) and Takumi Shigihara

Faculty of Education, Tokyo Gakugei University, 4-1-1 Nukuikita, Koganei,
184-8501 Tokyo, Japan
shumatsu0@gmail.com

Abstract. We investigate the effects of superimposing animated graphics of a virtual character (VC) on physics simulation learning material. Eye-tracking experiments revealed that the VC drew attention to the animated simulation display in a tutorial mode in which the user remained passive to instruction. No similar effect of superimposing was found when the user was engaged in the interface of waiting for key-in mode. Visual incongruity together with contextual congruity is believed to work as a stimulus to raise interest in the intuitive elements of the material. As one application, we describe the development of a visual annotation system based on augmented reality technology. The annotations were visualized in stereoscopic three-dimensional graphics using a see-through wearable binocular-type display. This system is useful for users to obtain in-depth knowledge individually from a large projected image shared with an entire class. This is expected to enable learners to retrieve knowledge at their own paces, while raising interest in the entire view.

Keywords: Learning materials · See-through wearable display · Augmented reality · Stereoscopic 3D

1 Introduction

Blackboard displays, projection screens, or large-format televisions are popular devices for beginning a lecture or sharing ideas among many participants. In addition to providing these popular methods of presentation, a projection mapping technology can afford an impressive visual experience [1]. In addition, augmented reality (AR) technology can annotate real objects or phenomena with virtual objects.

Santos et al. classified the learning use of AR into three categories [2]:

1. Real-world annotation.
2. Contextual visualization.
3. Vision-haptic visualization.

As an example of real-world annotation, AR technology was applied in web content to annotate the image of a real apparatus [3]. AR graphics of balloons and arrows located on images of buttons and switches provided explanations of their functions. As an example of visual-haptic teaching material, images of calculated force vectors were projected on a real experimental set in which three weights were

connected at a node and hanged with strings. In this learning material of force balance, one could sense the change of the reaction force by pulling a string connected to the node. The haptic sensing was expected to be reinforced by the AR vector images.

In many cases of AR presentation, viewers feel incongruity between virtual AR graphics and real objects. This incongruity may be attributable to an inconsistency in their contexts or visual perspectives [4]. Further, if such incongruity stimulates the human perception of images, virtual objects are expected to superimpose on real objects in order to raise the interestingness of a real scene.

For interactive material, a tutorial is often required to enable users perform a complete function of the material; particularly for self-learning, a tutorial must explain virtual experiments using various parameters. To stimulate user's curiosity regarding the content, a tutorial with a graphic that has paired conditions of visual incongruity and contextual congruity might be effective when superimposed on the content.

A virtual tutor has been reported to enhance interactivities between learners and online contents [5]. In this study, we consider a virtual character (VC) as a tutor and superimpose it on the material. The VC explains the parameter setting step by step, and users follow the steps by clicking at the indicated buttons. We compare the dwell times of participants' eye movements on various parts of the simulation content in a VC-led tutorial using a simple letter-prompting tutorial.

The flow of our tutorial is one way, with the user just following each instruction step. Similarly, the presentation of learning contents in the classroom is generally conducted with an instructor showing a scene and pointing out specific locations in the scene step by step.

In this study, a pilot system for a wearable see-through head mount display (HMD) is set up to obtain images that describe knowledge associated with the parts of a projected large image. AR is used to display the related image contents that are superimposed on the large image through the HMD. A superimposed AR image annotating the parts of a large image is expected to raise the interestingness of the entire scene as an effect of moderate incongruity. In addition, a stereoscopic three-dimensional (S3D) display is beneficial for educational purposes because of an enhancement of the recognition of three-dimensional (3D) structures [6]. In this study, the binocular-type see-through HMD enables the superimposition of S3D images on the main scene.

2 Method

2.1 Eye-Tracking Measurement

Learning Material. To detect the effect of a VC, we created learning material with a simple scenario of a virtual experiment on free-falling objects. In the experiment, the falling time of an object from a fixed height to the ground was measured. The purpose is to learn that the falling time is the same for objects of any mass, but the falling time itself depends on the mass of the planet. This material was developed using Unity [7], a 3D game development environment, with the free-fall motion of spherical objects simulated based on Unity's built-in physics calculation engine.

The three-dimensional VC selected for this study was “unity-chan!,” with digital image data provided and licensed for game developers by Unity Technologies Japan [8]. Voices of the character were composed using speech synthesis software, “Voiceroid + Yuzuki Yukari EX” [9]. The character’s waiting and jumping motions were built-in motions of the character’s package. The composition of additional motions for tutorial gestures was performed using the three-dimensional computer graphics animation software “MikuMikuDance” [10].

This VC behaves as a tutor in our material, by talking and responding with gestures to the user. Superimposing this VC on the free-fall simulation appears somehow incongruous, but the tutorial exhibits a contextual congruity with the simulation.

The material consists of two modes: “tutorial mode” and “waiting for key-in mode (or waiting mode, briefly).” In the tutorial mode, the aim and process of the virtual experiment are explained using text prompts and the VC’s speech and gestures. The tutorial tells the user to click buttons to set parameters and perform experiments. The participants act passively in the tutorial mode to the direction of the tutorial.

In the waiting for key-in mode, the application waits for users’ input. Users select buttons to click and perform the virtual experiment. The VC responds to the user’s button clicking, repeating the parameter values.

Eye Tracking. To detect the effect of displaying a VC, eye-tracking measurement was performed using “ViewTracker” (DITECT corporation), a head-mounted-type eye tracker. Images were shown to a participant in a 27-inch liquid crystal display; the participant was sitting approximately 60 cm from the display.

Depending on the presence or absence of the VC, viewing types were classified as follows.

- (vt1) tutorial mode without VC.
- (vt2) tutorial mode with VC.
- (vw1) waiting for key-in mode without VC.
- (vw2) waiting for key-in mode with VC.

Table 1 summarizes the viewing classification.

Table 1. Classification of four viewing types

Mode	View	Character display	Leading element	Feedback to user input
Tutorial	vt1	None	Prompt	Prompt & parameter display
	vt2	Tutor & response	Prompt & VC	Prompt, parameter, character
Waiting for key-in mode	vw1	None	None	Parameter display
	vw2	Response	None	Parameter, character action

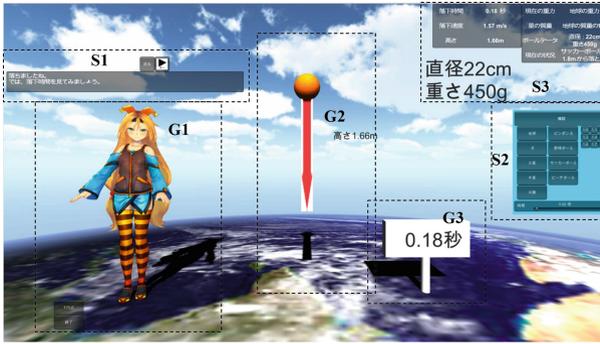


Fig. 1. Tutorial-mode image of the learning material. Broken lines show graphics and symbol component areas G1–S3, explained in Table 2.

In the eye-tracking experiments, we measure dwell times inside the component group areas. We define the dwell time as the time spent in a specific area of the interface. The interface of the material consists of graphic and symbolic component groups. The graphic component is intuitive, including the virtual tutor, free-fall animation, and display of the falling time. The latter two components are essential parts of this material. The symbolic component is comparatively logical, including the text prompt with a “next and back” button, buttons for selecting parameters, and display of parameter values. The component classification and the corresponding symbols of eye movement dwell time are summarized in Table 2.

Table 2. Interfacial components and the symbols of corresponding dwell times

Interface	Characteristics	Area of components	Symbol of dwell time	Symbol of area (Fig. 1)
Graphic	Intuitive, empirical, experimental	VC	T_{chr}	G1
		Free-fall animation	T_s	G2
		Display of time to fall	T_t	G3
Symbolic	Logical, procedural	Text prompt & next button	T_{pr}	S1
		Buttons for parameter selection	T_{bt}	S2
		Display of parameter value	T_p	S3
Outside		No component, outside the display	T'	

We compare the dwell time in the graphic interface with that in the symbolic interface, excluding from analysis the dwell time on the VC T_{chr} ; that is, the total time of observation T is calculated as $T = T_s + T_t + T_{pr} + T_{bt} + T_p + T'$. The ratio of the dwell time in the graphic interface to the total time of observation σ_{gr} is defined as

$$\sigma_{gr} = (T_s + T_t)/T \quad (1)$$

In addition, the ratio of the dwell time in the symbolic interface σ_{sb} is defined as

$$\sigma_{sb} = (T_{pr} + T_{bt} + T_p)/T \quad (2)$$

Because T includes the tracking time for looking outside the functional interfacial areas, the sum of the values σ_{gr} and σ_{sb} is $\sigma_{gr} + \sigma_{sb} \leq 1$.

Figure 2 shows examples of the four view types, indicating the interfacial areas from G1 to S3 in Table 2 for each view type. The free-falling objects are not shown in G2 in these figures. The boundary lines were drawn manually after each session, considering the measured eye movement trajectories.

S3D AR with HMD. To examine the superimposition of the S3D AR display through the see-through HMD, stereoscopic simulation material of molecular motions with Lennard-Jones and Coulomb potentials was created and presented using a dual projector with linear polarizers and projected on a polarization-preserving screen. A side-by-side S3D molecular motion graphic was created using Adobe Flash with Papervision3D and was made accessible on the web. AR markers were shown at the corners of the image.

An EPSON T-200 see-through HMD was used to present the graphics associated with the simulation content. A laptop equipped with a web camera was used to capture the AR markers and render the AR graphics, and the display was mirrored to the HMD. AR was created using Unity with a Vuforia [11] AR package. Only the AR graphics were rendered on the black background, without rendering a background video of the environment, enabling the user to superimpose the AR graphics at an arbitrary position in the projected image. The HMD eyeglasses were covered with polarization filters for viewing the S3D image on the screen.

3 Results and Discussion

3.1 Effect of VC on a User's View

Fifteen university students of the faculty of education participated in the eye-tracking measurements for the tutorial mode, and twelve or thirteen of them attempted it twice. Ten of the students participated also in the measurements for wait for key-in mode. The number of data and mean viewing time are summarized in Table 3. The viewing time of tutorial mode with VC was slightly longer than that without VC. This partially reflects the fact that participants looked at VC in addition to the text prompt from time to time. On the other hand, there was no significant difference between with and without VC in the waiting mode.

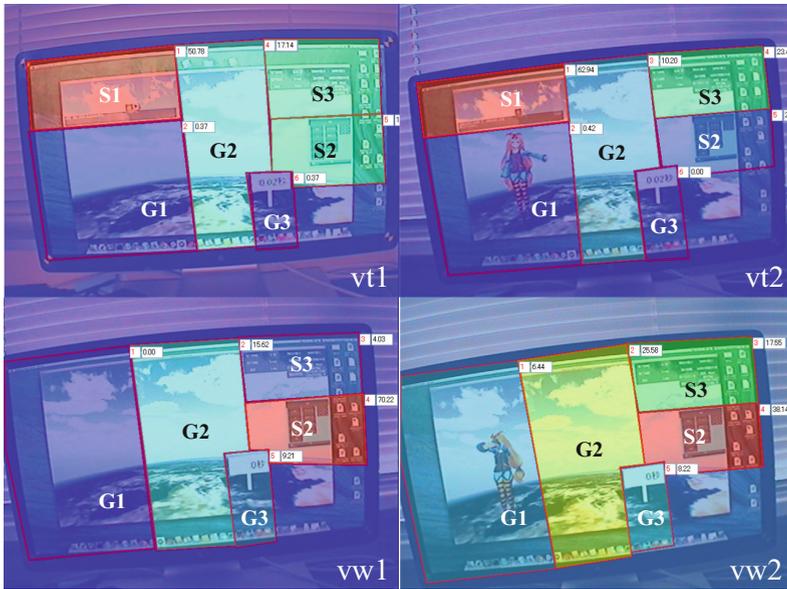


Fig. 2. Views of the learning material and the colored rectangular areas for detecting participant dwell time. Left top (vt1)-tutorial mode without VC, Right top (vt2)-tutorial mode with VC, Left bottom (vw1)-waiting mode without VC, and Right bottom (vw2)-waiting mode with VC.

Figure 3 shows the results of the mean dwell time of each component area for the tutorial mode. Total dwell times spent in the text prompt were 44 % (with VC) and 59 % (without VC). Total dwell times spent in the free-fall animation area were 22 % (with VC) and 15 % (without VC). The dwell time in the animation area was $T_s = 18.4 \pm 11.5$ s with VC and $T_s = 10.8 \pm 5.6$ s without VC. The equality of these mean values was rejected at a significance level of 0.05 (*) using the Welch t test. These results suggest that attention to the animation increased with animated VC.

Figure 4 shows the dwell times of wait for key-in mode with and without VC. Total dwell times spent at the control button area were 39 % (with VC) and 46 % (without VC). At the simulated animation area, total dwell time spent was 21 % (both with and without VC).

The display of falling time and parameters were regarded as feedback components. Attention to these feedback components increased from tutoring to wait mode. In

Table 3. Viewing time of eye-tracking measurements

Mode	View	Character display	Viewing time [s]	No. of samples
Tutorial	vt1	None	71.7 ± 20.1	27
	vt2	Tutor & response	82.6 ± 10.6	28
Waiting for key-in mode	vw1	None	72.4 ± 8.6	10
	vw2	Response	74.2 ± 4.4	10

particular, raised attention to the time display implies that interest in the quantitative results of this simulation increased more in the wait mode than in the tutorial mode. This suggests a change in participants’ intention from passively following the tutoring to actively experimenting to see the results of falling time.

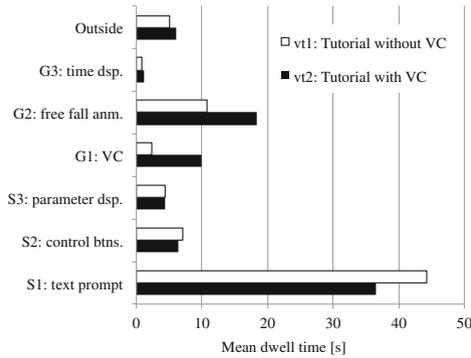


Fig. 3. Mean dwell times for graphic and symbolic components in the tutorial mode

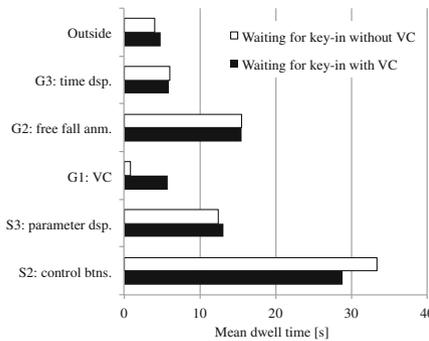


Fig. 4. Mean dwell times for graphic and symbolic components in the waiting mode

Figure 5 shows the plots of σ_{gr} and σ_{sb} of the tutorial mode with and without VC. As an overall tendency, participants paid more attention to the symbolic components than the graphic components. However, relatively higher values of σ_{gr} were found with the presence of VC. Specifically, mean values of σ_{gr} were $\sigma_{gr} = 0.26 \pm 0.16$ and $\sigma_{gr} = 0.18 \pm 0.10$ with and without VC, respectively. The Welch t test showed that the p value of mean σ_{gr} was 0.055, indicating a rejection of equality of σ_{gr} at a significance level slightly greater than 0.05 (*). Mean values of σ_{sb} were $\sigma_{sb} = 0.65 \pm 0.19$ and $\sigma_{sb} = 0.72 \pm 0.14$ with and without VC, respectively. The equality of these values was not rejected at the same significance level. Although the text prompt had a strong attention-drawing effect, the results suggest that the presence of VC induces attention to the other graphical components, even when the user is in a reactive state.

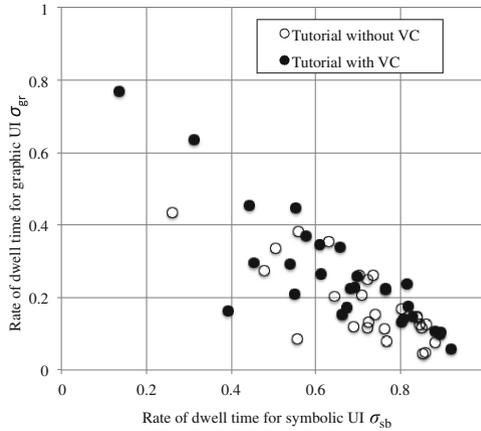


Fig. 5. Plots of the rate of dwell time for graphic and symbolic components in the tutorial mode.

Figure 6 shows the plots of σ_{gr} and σ_{sb} of the waiting VC with and without VC. Mean values of σ_{gr} and σ_{sb} were $\sigma_{gr} = 0.32 \pm 0.12$ and $\sigma_{gr} = 0.30 \pm 0.08$ with and without VC, respectively, and $\sigma_{sb} = 0.61 \pm 0.32$ and $\sigma_{sb} = 0.64 \pm 0.30$ with and without VC, respectively. By the Welch t test, the p values of σ_{gr} and σ_{sb} were 0.67 and 0.72, respectively. As a result, the equality of the mean values of σ_{gr} and σ_{sb} with and without VC were not rejected by the Welch t test. Thus, in the waiting mode, in which participants use the symbolic interface at will, no particular effect of VC on dwell time was detected.

3.2 Presentation of S3D AR Images by HMD

Based on the above results, we developed a presentation method for providing images to be superimposed on a large image shared in the classroom. Side-by-side S3D images were projected from a polarized dual projector on a polarization-preserving screen. Participants viewed this projected image through the see-through HMD in which left and right eyeglasses were covered with polarizers.

The projected image included AR markers, as shown in Fig. 7 left. Positions of AR markers in the left and right images of the side-by-side display were equivalent, resulting in the superimposed marker images having no positional displacement. Consequently, the AR marker depth positions in the S3D display were on the plane of the screen.

The AR application for detecting markers and generating side-by-side stereoscopic AR images was run using a laptop with a web camera. The user obtained the images of AR markers using the web camera. AirPlay mirrored the generated AR images to the see-through HMD in side-by-side 3D mode. In the AR display, a created image targeted by the marker was rendered with a black background, as shown in Fig. 7 right. The exclusion of environmental video enabled the superimposition of a single AR image on the projected image on the screen.

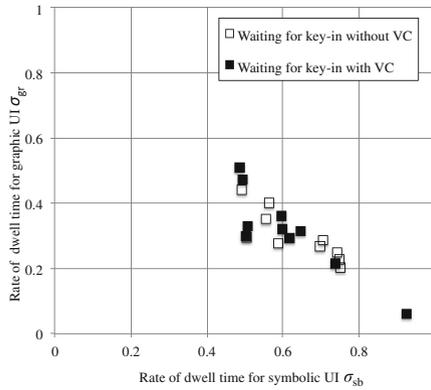


Fig. 6. Plots of the rate of dwell time for graphic and symbolic components in the wait for key-in mode.

Because the web camera was connected to a laptop, users could adjust the position of the AR image in two ways. First, by moving the web camera, one could change the position of the AR image in the HMD-view frame. Second, the user's head motion could change the position of the AR image in relation to the screen image.

In this method, the user uses the see-through HMD as a viewer of the laptop. PC operation wearing the HMD is possible if the HMD-vision background color is sufficiently dark. In our method, obtaining an AR image from the screen image is one of the activities enabled by combining HMD and PC. The web camera is used for AR image switching.

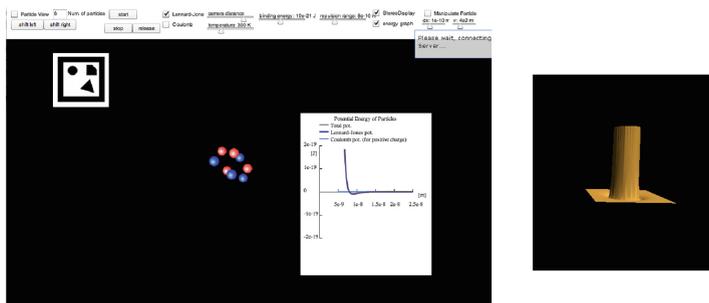


Fig. 7. Screenshot of molecular motion content. Left image: Interacting particles in the central area are displayed stereoscopically on the screen. The graph inside shows the Lennard-Jones potential energy curve. The AR marker is in the left up position. Right image: A 3D graphic of the Lennard-Jones potential superimposed as an S3D image on the screen projection of the left image through the see-through HMD view.

4 Conclusion

We found that users in a passive state were more aware of essential graphic components with the presence of the VC. Such a VC effect was not detected when the interface was in waiting for key-in mode and the users were active in operating it. It was suggested that a VC acted as a stimulus for engaging users in the essential part of the learning material, rather than simply following a procedure. Relevant superimposition of such salient graphics as a visual stimulus to the learning of materials opens the possibility for a user to be active. This is applicable to the situation of classroom learning, in which an instructor explains a large image projection to the entire group of participants.

To apply this activation effect in classroom instruction, a binocular-type see-through wearable display was introduced to represent stereoscopic 3D graphics mirrored from individuals' laptops. The graphics were generated as AR images by detecting markers attached to the large image. The AR technology played a role in annotating the shared image, while it provided stimuli to activate personal learning. Using this method, we expect to enable learners to retrieve knowledge at their own paces, while raising interest in the entire view.

Acknowledgments. A part of this study has been funded by a Grant-in-Aid for Scientific Research (C) 21500842 from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The authors would like to thank Enago (www.enago.jp) for the English language review.

References

1. Kanazawa, A., Asai, T., Minazuki, A., Hayashi, H.: IIAI 3rd International Conference on Advanced Applied Informatics, pp. 247–252 (2014)
2. Santos, M.E.C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., Kato, H.: *IEEE Trans. on Learn. Technol.* **7**, 38–56 (2014)
3. Tsuchida, S., Yumoto, N., Matsuura, S.: Development of augmented reality teaching materials with projection mapping on real experimental settings. In: Stephanidis, C. (ed.) *HCI 2014, Part II. CCIS*, vol. 435, pp. 177–182. Springer, Heidelberg (2014)
4. Liestøl, G., Morrison, A.: *IEEE International Symposium on Mixed and Augmented Reality—Arts, Media, and Humanities*, pp. 23–28 (2013)
5. Yamamoto, T., Miyashita, R.: *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, pp. 564–568 (2008)
6. Yoshii, T., Matsuura, S.: *Proceedings of the International Conference on 3D Imaging* (2011). doi:[10.1109/IC3D.2011.6584372](https://doi.org/10.1109/IC3D.2011.6584372)
7. Kim, S.L., Suk, H.J., Kang, J.H., Jung, J.M., Laine, T.H., Westlin, J.: *IEEE World Forum on Internet of Things*, pp. 21–26 (2014)
8. Unity-chan. <http://unity-chan.com>. Accessed 20 February 2015
9. Kenmochi, H.: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5385–5388 (2012)
10. Xu, J., Takagi, K., Sakazawa, S.: *Proceeding of SIGGRAPH Asia* (2011). doi:[10.1145/2073304.2073332](https://doi.org/10.1145/2073304.2073332)
11. Qualcomm, Vuforia. <https://www.qualcomm.com/products/vuforia>. Accessed 20 February 2015