



AR Cooking: Comparing Display Methods for the Instructions of Cookwares on AR Goggles

Hiroki Hasada^(✉), Junjian Zhang^(✉), Kenta Yamamoto^(✉),
Bektur Ryskeldiev^(✉), and Yoichi Ochiai^(✉)

Digital Nature Group, Pixie Dust Technologies Inc.,
University of Tsukuba, Tsukuba, Japan

{hirokihasada, tyookk, kenta.yam, bektour}@digitalnature.slis.tsukuba.ac.jp,
wizard@slis.tsukuba.ac.jp

Abstract. We explore effective display methods for instructions on complex tools on augmented reality (AR) goggles. We focused on three types of cookwares as complex tools in everyday life: avocado cutter, orange opener, and can opener. We implemented three AR applications for displaying the usage of three cookwares: images with text, videos, and three-dimensional (3D) animation. In this study, we investigated how each of these AR applications affects the understanding of complicated products and their applications. Using Microsoft HoloLens, we conducted user studies with the three aforementioned display methods. We measured the task completion time and investigated the psychological element. Based on the task completion time and whether the task could be completed, we found that the most efficient display method for the instruction of cookwares on AR goggles is 3D animation. We investigated how participants felt about cooking using AR goggles based on participant's comments. We believe that our contribution can be useful for the design of AR-based application.

Keywords: UX and usability ·
Qualitative and quantitative measurement and evaluation ·
User experience

1 Introduction

AR goggles are considered effective for various instructions such as cooking. In recent years, many cooking assistant technologies have been studied. Assistance cooking is an important research field for enriching human life. When people are cooking, one may encounter a cookware whose usage is not trivial by appearance. It is particularly typical in novelty products such as avocado cutter, orange opener, and can opener in Fig. 2(a)–(c). Currently, tutorials involving images with text and videos are typical as a method for presenting the usage of such products. In traditional cooking scenarios, people are used to learning cooking

Table 1. Pros and cons of the three kinds of the media for instruction.

	Pros	Cons	Example
Paper manual	Almost all people are used to it as traditional media	More maintenance cost than digital data	Cook books Assembly manual
Tablet device	Rich contents can be achieved through Internet. Digital data is easily reserved	Hard to operate when hands are wet. Cannot operate without electric power	iPhone ^a iPad ^b
AR goggles	Understanding the needed information better by adding extra virtual information into real world	Current devices are heavy. The viewing angle is narrow. Cannot operate without electric power	Microsoft HoloLens ^c Magic Leap ^d

^a <https://www.apple.com/jp/iphone/> (Last accessed 2019-02-15)

^b <https://www.apple.com/jp/ipad/> (Last accessed 2019-02-15)

^c <https://www.microsoft.com/ja-jp/hololens> (Last accessed 2019-02-15)

^d <https://www.magicleap.com/> (Last accessed 2019-02-15)

by reading the instructions through paper media or tablet devices. However, both hands may get wet or dirty during cooking, which can be a psychological barrier to touching paper or tablet devices directly. In addition, most tablet devices may be damaged when they are splashed with water. AR technologies are thought to be an effective approach to such problems in these traditional media. Table 1 shows the advantages and disadvantages of the three media. In this study, we focus on exploring an effective display method for cooking on Microsoft HoloLens, an AR goggles. We conducted experiments on cookwares as shown in Fig. 1(a). In this experiment, we displayed images with text, videos and 3D animation as shown in Fig. 1(b)–(d), on Microsoft HoloLens. Quantitative and qualitative results are presented from user surveys on 35 different participants.

The contributions of this research are as follows:

- We investigated how three display methods with Microsoft HoloLens affect users on the efficiency of cooking tasks.
- We discuss the evaluation results by quantitative and qualitative methods.
- We discuss the user experience from at free-description-type questionnaire.

2 Related Work

2.1 Instruction Method by AR Technology

Currently, many practical applications of AR technology are considered, e.g. medical and assembly work. Furthermore, research to compare AR goggles with

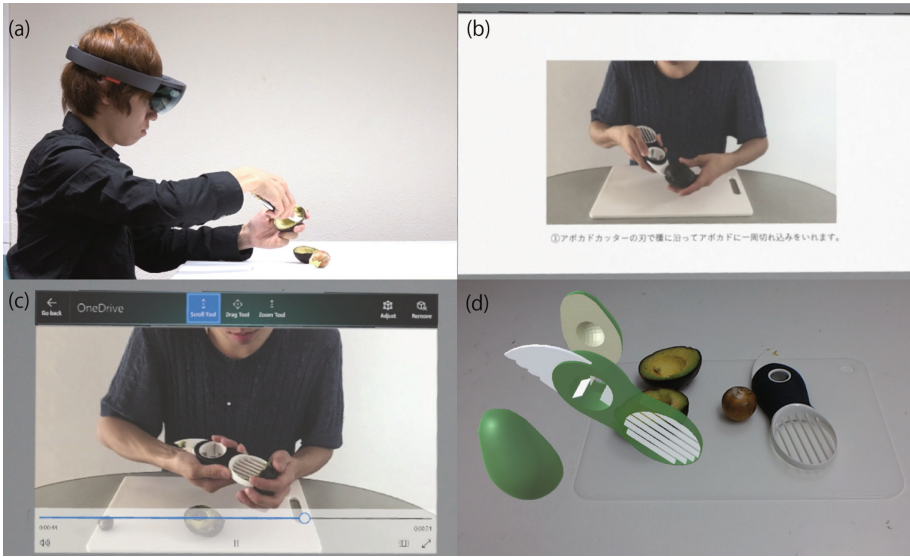


Fig. 1. Variety of cooking instructions. (a) the participants completed the task while watching the 3 instructions displayed on the AR goggles, (b) images with text, (c) the video, (d) the 3D animation.



Fig. 2. Variety of cookwares. (a) the avocado cutter, (b) the orange opener, (c) the can opener.

various display media have been conducted [1, 3, 4, 8, 11, 13, 15]. In 2017, Orsini et al. published a practical application using Microsoft HoloLens on cooking.¹ However, the relationship between cooking and display method was not explored. Furthermore, they did not explore what was effective for displaying on AR goggles. Cooking is a typical act performed in many families. Exploring the relationship between the display method and users is an important study related to the sense of use when making practical applications. Therefore, we attempted to explore the relationship between cooking and display methods on AR goggle and used it for future practical application designs.

¹ <http://www.ece.rutgers.edu/sites/default/files/capstone/capstone2017/posters/S17-35-poster.pdf> (Last accessed 2019-02-15).

2.2 Cooking Support System

Research on cooking assistance with a tablet device has been discussed in [5, 14]. For a tablet device, a location must be secured to place it in. In addition, we cannot operate tablet devices when they are wet. Furthermore, while cooking, it is inevitable for the hand to become wet or dirty. However, these devices are difficult to operate in such a case. In addition, because the image size that can be displayed depends on the device size, it is difficult to observe presbyopia or a small screen for people with poor vision. Many methods using a projector are being studied as a method that does not require touching by hand [2, 6, 7, 9, 10, 12]. When projecting with a projector, if an object exists between the projector and the projection plane, a person who cooks is hard to see because a shadow has been generated. Depending on the installed position and orientation of the projector, the plane that can be displayed is limited. In addition, depending on the kitchen, a space for installing the projector might not be available. Moreover, depending on the kitchen, it is conceivable that an ideal plane that can be projected is non-existent. In the case of AR goggles, it is not necessary to project the recipe on the plane because the recipe can be arranged anywhere within the space; therefore, the projection space is not an issue. Orsini et al. proposed a system for displaying recipes using object recognition on AR goggles. In this system, three-dimensional (3D) animation was used to display, and when it becomes difficult to convey, the recipe of the animation can be viewed. However, they did not mention which display method was easier to convey. Various approaches have been reported for cooking-assistance systems. However, experiments focusing on display methods on AR goggles have not been conducted.

3 Design of Displaying Instructions on AR Goggles

Figure 3 shows overview of the experiment. Images with text and videos are displayed so that they are projected on a whiteboard about 1 m away from AR goggles. 3D animation was displayed in midair about 30 cm away from AR goggles. 3D animation can be seen simultaneously with the cooking area. However, images with text and videos cannot be seen simultaneously with the cooking area.

3.1 Images with Text

Figure 4 shows instruction of images with text of avocado cutter. Currently, the most popular instruction platform is the paper manual. Images with text are created by referencing the paper manual. Images were created by capturing screenshots from videos shown in Fig. 5. One should click on the right side of the screen to advance to the next page and click on the left side to return to the previous page. The image was fixed on a whiteboard at a distance of approximately 1 m and projected such that the entire image would fit in the viewing angle of Microsoft HoloLens to the participants. If we display the images

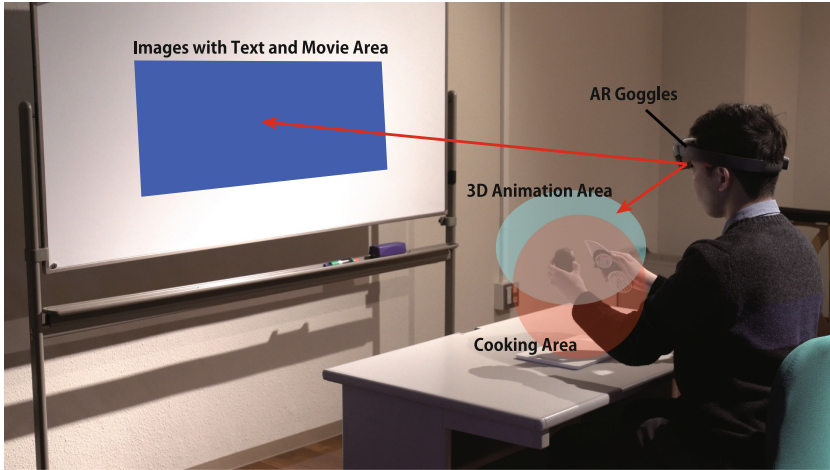


Fig. 3. Overview of experiment. Images with text and videos are displayed so that they were projected on a whiteboard about 1 m away from AR goggles. 3D animation was displayed in midair about 30 cm away from AR goggles.

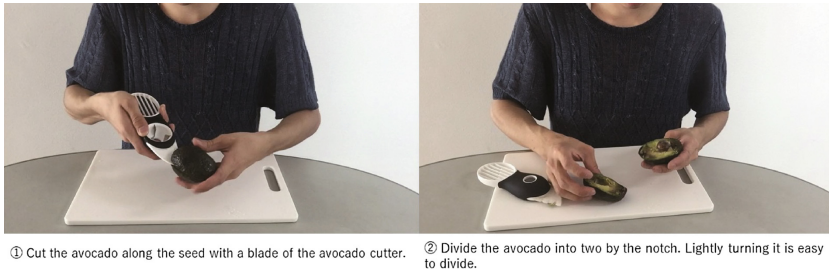


Fig. 4. Images with text of the avocado cutter. Images were created by taking screenshots from videos in Fig. 5. Click on the right side of the screen to advance to the next page and click on the left side to return to the previous page.

in front of the participant, they can not see their hands. Therefore, we display the images on the whiteboard in front of the participants. We add text to each image because there is no visual movement information. This display method was considered to be inferior to the video and 3D animation in that there is no visual information of movement.

3.2 Video

Figure 5 shows instruction of the video of avocado cutter. The audio of the video was turned off. We can return to the video using the control bar, and we can stop the video with the play/pause button. The video was fixed on a whiteboard at a distance of approximately 1 m and projected such that the entire video would

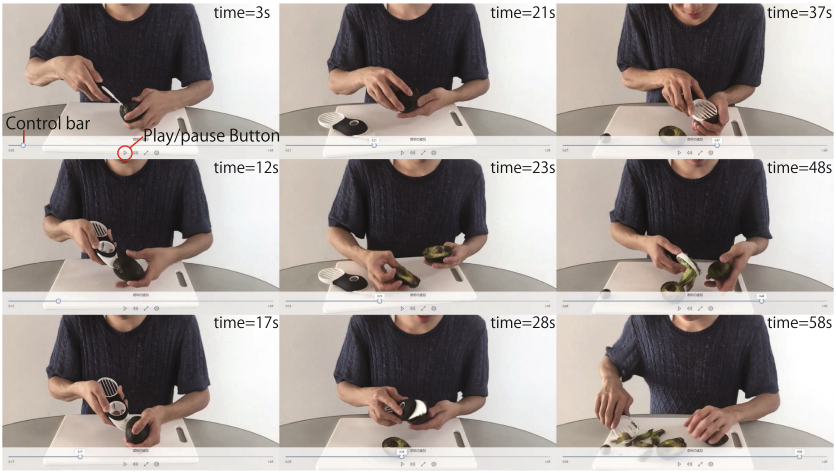


Fig. 5. The video of avocado cutter. We took the video of peeling avocado using avocado cutter from the front. We can return the video using the control bar, and we can stop the video with the play/pause button.

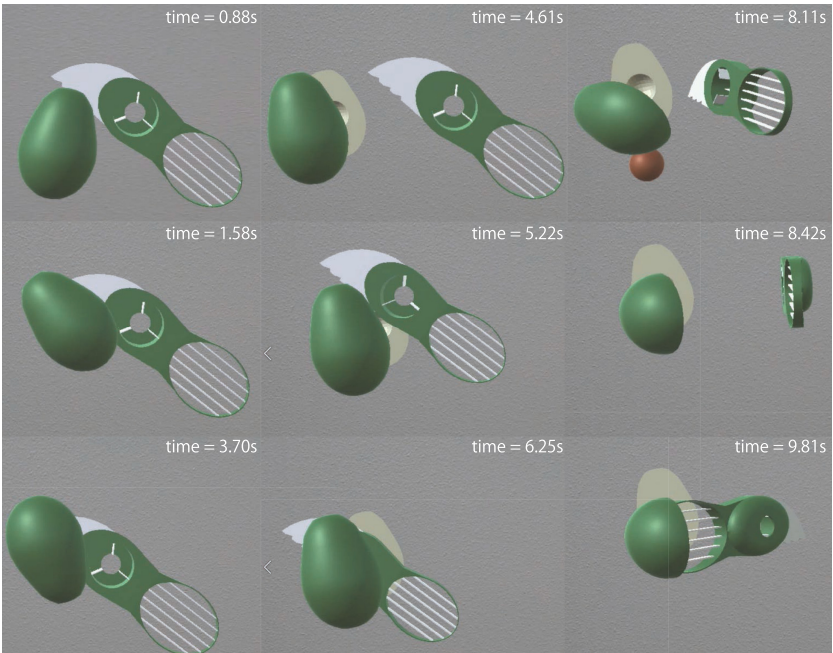


Fig. 6. 3D animation of avocado cutter. 3D animation was created to loop automatically.

fit in the viewing angle of Microsoft HoloLens to the participants. If we display the video in front of the participant, they can not see their hands. Therefore, we display on the whiteboard in front of the participant. We did not instruct by voice or subtitle. Therefore, we verified whether instructions are conveyed only with the video. We considered that this display method was superior to images with text because the visual movement information can be conveyed. However, we considered that this display method was inferior to 3D animation because 3D information cannot be conveyed and we cannot display at hand.

3.3 3D Animation

Figure 6 shows instruction of 3D animation of avocado cutter. The 3D animation instruction was implemented by Unity², and displayed at a distance of approximately 30 cm from the participant. The 3D animation was created to loop automatically. We display the 3D animation in front of the participants because their hands are not obscured by images or videos. We considered that participants prefer 3D animation to images with text and video because they could observe the three-dimensional information, as well as see both hands and instructions at the same time.

4 User Study

4.1 Methodology

Study Process. The independent variables were the display method and cookwares. The dependent variables were seven-point Likert scale scores for measuring the efficiency and task completion times. Participants practised the hand gesture of Microsoft HoloLens before doing the task. They learned the air tap (corresponding to click on PC) and tap & hold (corresponding to drag & drop on PC). Task completion time was measured when each work was finished. The avocado-cutting task was deemed completed when the last slice was cut. The orange-opening task was deemed completed when the orange was completely peeled. The can-opening task was deemed completed when the participants emptied the can. We chose the task completion time as a quantitative evaluation. After the tasks were completed, the participant was asked to complete post-test questionnaires: qualitative questionnaire, the demographic information and description formula.

Participants. Thirty-five participants (18 women and 17 men, with six left-handed participants) of ages from 18 to 24 participated ($M = 20.8$, $SD = 1.4$) in the experiments. Some of the participants performed multiple tasks using the two or three cookwares, and 68 trials were conducted. A previous questionnaire is conducted to ensure that all the participants have never use the cookware ever before and were not aware of the specific usage of the tool before the experiment.

² <https://unity3d.com/jp> (Last accessed 2019-02-15).

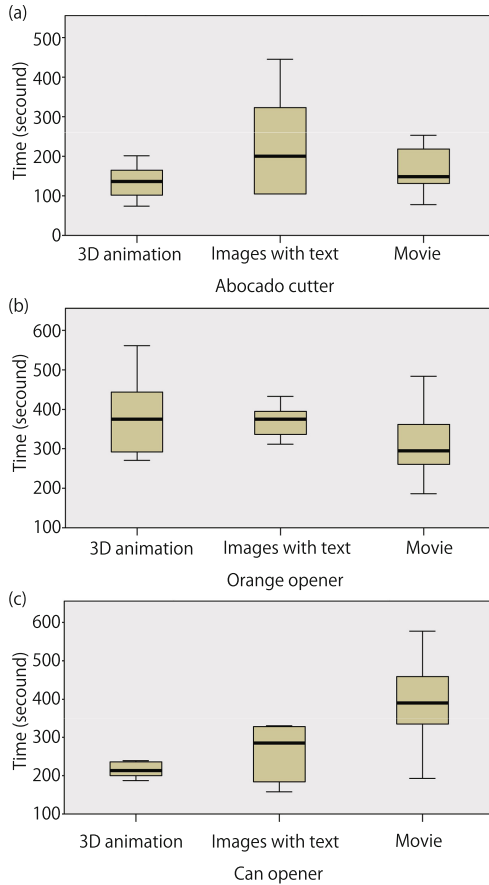


Fig. 7. Task Completion Times of three cookwares. (a) shows the result of the avocado cutter. (b) shows the result of the orange opener. (c) shows the result of the can opener.

It is noteworthy that all participants used the cookware for the first time in the experiment. It is necessary for each participant to not know the cookware usage presented to them. Therefore, cases exist where a few participants did not know the usage of one or two tools among the three tools. The number of people who did not know the usage of only one tool was 13; the number of people who did not know the usage of two tools was 11; and the number of people who did not know the usage of all three tools was 11.

4.2 Results

Task Completion Times. One participant quit the experiment for the task of images with text for a can opener. However, all other participants completed all the tasks. Figure 7(a)–(c) shows the task completion time result of each cookware. After removing the data points of participants who quit from the 68 data

points, we excluded data that are over $+3.0SD$ from the time required for each task as outliers. One data point was excluded from 67 data points. A one-way ANOVA was conducted to study the effect of the instructional method on the task completion times for each cookware. When the ANOVA indicated a significant difference between systems, pairwise comparisons were conducted using the Bonferroni correction. The effect was statistically significant, i.e., $F(2, 18) = 6.67$, $p = 0.036$. In the can-opening task, the ANOVA analysis indicates statistically significant effects between the video and 3D animation ($p = 0.036$). There is no significant effect between images with text and the video ($p = 1.000$), and between images with text and 3D animation ($p = 0.203$). However, no statistically significant difference is shown in any display method for the avocado cutter and orange opener.

Questionnaire Results from Participants. We conducted a questionnaire survey for five items for qualitative evaluation. The results of the questionnaire are reported in Fig. 8(a)–(e). The participants answered the questionnaires on items of efficiency, easiness, pleasantness, satisfaction, and hardship on a seven-point Likert scale and described the reasons for each of their responses. We conducted the ANOVA test on the questionnaire with the Likert scale. However, no significant difference was shown from any of the questionnaires. Therefore, we decided to explore why participants experienced efficiency or understandability from the description of the reasons for the five questionnaires. Sixty comments were obtained for each of the five questionnaires, except for no answers for 68 trials, and approximately 300 total descriptions were obtained.

5 Discussion

This section discusses the experimental findings based on the stated hypotheses. The implications of the results on the theoretical model are investigated, and further insights into the influence of AR in human performance and perception are provided.

5.1 Task Completion Times

From the results of the experiments on the cost time, the 3D animation approach was significantly faster than the video in the can-opening task. No significant difference is found in the avocado-cutting and orange-opening tasks. We consider that the effective display method depends on the task. However, depending on the task, it was shown that the task completion time by the display of 3D animation can be shorter than by the display of video. We consider that this result has been caused because the operation gesture is difficult for participants who are new to Microsoft HoloLens. In the display of the video, the participants must use the air tap gesture to play/pause a button and tap & hold of the control bar. On the other hand, the 3D animation is looped automatically by default. The control bar may be effective when watching a video on a smartphone or the

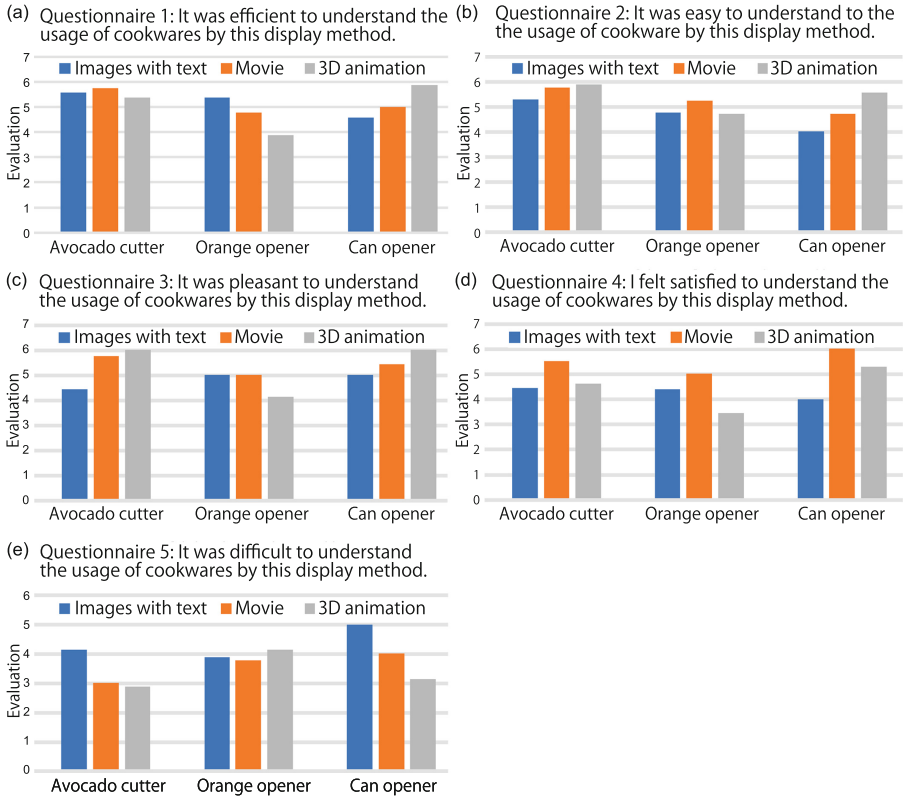


Fig. 8. Result of five Questionnaire. (a) Questionnaire 1: efficiency. (b) Questionnaire 2: easiness. (c) Questionnaire 3: pleasantness. (d) Questionnaire 4: satisfaction. (e) Questionnaire 5: hardship.

like. However, other operation methods may be better when watching a video on Microsoft HoloLens.

5.2 Participants' Comments

Sixty comments on each questionnaire were obtained except for no answer. We described the comments and their considerations in the following:

Questionnaire 1: Efficiency. Eleven people commented the following: “Microsoft HoloLens is usable even when I’m cooking with dirty hands.” Therefore, we consider that it is effective for participants to cook with AR goggles. As for the display by 3D animation, a participant commented “It is stereoscopic and easy to understand.” This indicates the possibility of 3D animation on AR goggles as for cooking. Meanwhile, 11 participants commented that there is little difference between AR goggles and other devices such as projectors, personal computers,

displays, and tablet devices as the reason for not understanding efficiency. Two of them complied with the images with text, six of them complied with the video and three of them with 3D animation. Therefore, we consider that the video is harder to understand the merit of AR goggles than other display methods.

Questionnaire 2: Easiness. We did not observe many biased comments. However, some of the participants who saw the 3D animation commented that its understanding was difficult owing to the lack of a hand model. Whether a hand exists in 3D animation will be addressed in future work. A participant who performed the avocado-cutting task using 3D animation replied “It is easy because I could mimic the movement of the 3D animation.” A load of work perception should decrease because it is sufficient to mimic the motion information. Furthermore, 3D animation can overlay on objects and ease objects observation, unlike the video. This is considered that the 3D animation is superior to the video as for the mimic of motion.

Questionnaire 3: Pleasantness. Eighteen participants commented “It was refreshing to understand how to cook on AR goggles.” We considered that they felt pleasant cooking on AR goggles because almost all of them were not used to AR goggles. As for the display of the 3D animation, one participant that they tried the orange-opening task commented “I am glad to see the backside of the task.” The video can convey only information from the front. On the other hand, the 3D animation can convey the information on the backside. Currently, as for our 3D animation, we can see the backside by changing the viewpoint. It is a future work to make the direction of 3D animation changeable by voice and hand gesture.

Questionnaire 4: Satisfaction. Regardless of the display method is used, the participants were satisfied when the task is completed. One participant who watched the 3D animation answered, “I felt satisfied because it was interesting to overlay an avocado on 3D animation.” It is difficult to overlay the avocado on the video or images with text. However, 3D animation can overlay the real object. Therefore, we consider that 3D animation is superior to the video and images with text. Further, one participant looking at the three types of display methods comments “Because the task difficulty level was not extremely high.” A different result may be obtained if the task becomes more difficult such as making a dish from a recipe.

Questionnaire 5: Hardship. Further, 14 reasons were related to Microsoft HoloLens as an AR device, such as “heavy,” “worrisome,” “difficult to operate,” “machine unfamiliarity.” These improvements can be expected by improving the performance of the device. As for AR goggles, the development of devices is remarkable. Therefore, we consider that it is necessary to perform new experiments every time a new device is launched.

5.3 Participants Who Quit

As for the can-opening task, one participant that see images with text stopped doing before they have been completed. Unlike the video and 3D animation, it is considered that images with text have little information on the movement of the can opener. Therefore, it is difficult to understand how to move the can opener from the instruction of images with text. We consider that participants need motion information to understand how to use.

5.4 Experiment Design

We consider that increasing the number of tasks can be effective in verifying the effectiveness of display methods on AR goggles. From performing this experiment by many tasks in the future, it is important to classify tasks according to effective display method when an effective display method differs depending on tasks. Based on the results, it is considered that by examining the characteristics of the tasks, effective display methods can be predicted for tasks that have not yet been verified. In this study, we conducted experiments on cookwares. However, we would like to investigate other tasks in the future. Participants were supposed to live by themselves and assumed experiments with participants aged 18 to 24, assuming university students to start cooking. Further, experiments involving homemakers who are more likely to cook can be considered in the future.

6 Conclusion

From the following two points, we considered that it would be better to use 3D Animation because:

- Task completion time of can opener with the 3D animation was significantly shorter than with the video.
- In the display by images with text, a participant could not proceed because information regarding the move and direction are not conveyed.

Although no significant difference was observed from the qualitative questionnaire, many positive opinions were obtained on using the AR goggles for cooking from the participant's description. Therefore, we considered using AR goggles for cooking to be sufficiently useful. However, we discovered that the current device presented problems from the aspects of weight, fitting feeling, viewing angle, and operability. Freshness and device problems for AR goggles may be worth investigating to study how they change over time as the user's experience with this technology accumulates. We believe that our research contributes to the field of future AR instructions.

References

1. Blattgerste, J., Strenge, B., Renner, P., Pfeiffer, T., Essig, K.: Comparing conventional and augmented reality instructions for manual assembly tasks. In: Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 2017, pp. 75–82. ACM, New York (2017). <https://doi.org/10.1145/3056540.3056547>
2. Bonanni, L., Lee, C.H., Selker, T.: Attention-based design of augmented reality interfaces. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005, pp. 1228–1231. ACM, New York (2005). <https://doi.org/10.1145/1056808.1056883>
3. Boud, A.C., Haniff, D.J., Baber, C., Steiner, S.: Virtual reality and augmented reality as a training tool for assembly tasks. In: IV, p. 32. IEEE (1999)
4. Caudell, T.P., Mizell, D.W.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. 2, pp. 659–669. IEEE (1992)
5. Hamada, R., Okabe, J., Ide, I., Satoh, S., Sakai, S., Tanaka, H.: Cooking Navi: assistant for daily cooking in kitchen. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA 2005, pp. 371–374. ACM, New York (2005). <https://doi.org/10.1145/1101149.1101228>
6. Ju, W., Hurwitz, R., Judd, T., Lee, B.: Counteractive: an interactive cookbook for the kitchen counter. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2001, pp. 269–270. ACM, New York (2001). <https://doi.org/10.1145/634067.634227>
7. Lee, C.H.J., Bonanni, L., Espinosa, J.H., Lieberman, H., Selker, T.: Augmenting kitchen appliances with a shared context using knowledge about daily events. In: Proceedings of the 11th International Conference on Intelligent User Interfaces, IUI 2006, pp. 348–350. ACM, New York (2006). <https://doi.org/10.1145/1111449.1111533>
8. Nilsson, S., Johansson, B.: Acceptance of augmented reality instructions in a real work setting. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2008, pp. 2025–2032. ACM, New York (2008). <https://doi.org/10.1145/1358628.1358633>
9. Sato, A., Watanabe, K., Rekimoto, J.: Shadow cooking: situated guidance for a fluid cooking experience. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014. LNCS, vol. 8515, pp. 558–566. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07446-7_54
10. Scheible, J., et al.: SMARTKITCHEN media enhanced cooking environment. In: Proceedings of the 6th International Conference on the Internet of Things, IoT 2016, pp. 169–170. ACM, New York (2016). <https://doi.org/10.1145/2991561.2998471>
11. Schwald, B., De Laval, B., Optronique Sa, T., Guynemer, R.: An augmented reality system for training and assistance to maintenance in the industrial context. *J. WSCG* **11** (2003)
12. Suzuki, Y., Morioka, S., Ueda, H.: Cooking support with information projection onto ingredient. In: Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction, APCHI 2012, pp. 193–198. ACM, New York (2012). <https://doi.org/10.1145/2350046.2350084>

13. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 73–80. ACM, New York (2003). <https://doi.org/10.1145/642611.642626>
14. Vildjiounaite, E., et al.: Designing socially acceptable multimodal interaction in cooking assistants. In: Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI 2011, pp. 415–418. ACM, New York (2011). <https://doi.org/10.1145/1943403.1943479>
15. Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C.: An augmented reality training platform for assembly and maintenance skills. *Rob. Auton. Syst.* **61**(4), 398–403 (2013)