A Psychometric Evaluation of the User Interaction System Combining an Aerial Display with Gesture Recognition

Isamu Nakao†, Shunsuke Takahashi†, Hiroshi Kobayashi† and Masahiro Yamaguchi†

Abstract A novel user interaction system combining an aerial display and a gesture interaction is expected to provide an excellent user experience. However, its psychometric effectiveness has not been sufficiently explored. In this reported study, we performed a usability test of the system that comprised an aerial display, gesture interaction, and suitable software content. Using a statistical analysis employing the Kruskal Wallis test and a factor analysis of the experimental results, the importance of a suitable combination of aerial display and gesture interaction as well as the novelty of the aerial display were demonstrated. Following this, a two-dimensional aerial display was considered to be adequately useful for a typical spatial user interaction system.

Keywords: aerial display, gesture recognition, user interaction, psychometric evaluation.

1. Introduction

Over the past decade, aerial displays that form a two-dimensional (2D) real image in mid-air between a display device and an observer have been proposed. User interactions that combine aerial displays and gesture recognition have also been attempted. In principle, an image can be formed simply using a large aperture lens in an aerial display by arranging the user’s observation conditions. Moreover, some innovative optical systems and devices for aerial display interfaces have been introduced. The hand gesture interactions have been significantly improved employing images acquired by cameras or new devices. User interaction such as touching a popped up aerial display seems to be a natural and acceptable approach for users. In this type of display system, the images are located closer to the user than the housing, although the displayed images do not have three-dimensional (3D) structures and nothing protrudes from the 2D display plane into space. From the user’s perspective, a question arises. In this situation, does the aerial display or the quality of gesture interaction contribute more to user interaction?

We focused on obtaining a fundamental answer to this question in the selection of the systems as well as in the quality of the user experience in the combined system.

Application targets were established based on off-the-shelf consumer electronics, mainly in the area of entertainment. This user experience depends on both the software content and the hardware configuration. We prepared a laboratory level aerial display apparatus, gesture recognition system, and appropriated software contents and performed subjective usability evaluations with test subjects. The questionnaire results were statistically analyzed and reported.

2. Experimental prototype user interaction system with aerial display

The hardware setup and the software content for the prototype user interaction system with an aerial display were prepared to execute a usability test employing select subjects as follows. A depiction of the whole setup is shown in Fig. 1. The setup comprised the housing and the control PC.

2.1 Hardware setup

An optical imaging setup is shown in Figs. 2A to C. In the aerial display configuration (Fig. 2A and B), the real image of a liquid crystal display (LCD) was formed using a Fresnel lens. A Fresnel lens with a 200 mm x 200 mm rectangular aperture was used. The focal length was 200 mm, and the spherical aberration was corrected. The optical path was reflected once before the lens so that the whole optical system could be installed in a small housing. At the aperture of the housing, transparent glass is situated and the images which were the aerial display were formed above the glass surface. A camera
Logicool C910 HD camera, Logitech, Lausanne Switzerland) used to take images of the user’s hands was installed at the edge of the table top aperture that was furthest from the user. The pictures acquired by this camera were used for gesture recognition. From the object shape, pattern, and colors, the user’s hands or fingers were recognized. The resulting gestures were translated to operation commands from the positions and the motions of the user’s digits and hands. This camera gesture recognition methods recognized the gestures not only in aerial display plane but also in any depth position. In order to activate this gesture recognition system only around the aerial display plane, two infrared (IR) sensors were installed as shown in Fig.1 to detect the presence of the user’s hands or fingers around the aerial display plane. Trigger signals for activations were generated and used in the following manner. The analog output signals from the IR sensors were converted to digital signals and were transferred to the control PC. The PC software was programed so that the gesture recognition system operated only in case both sensor’s signals indicated the presence of the digits and the hands simultaneously. Based on these functions, the setup can be classified into three types depending on the purpose of the experiment. Each feature of the three types of structures A to C is as follows:

A) higher aerial display quality and moderate gesture interaction quality;
B) lower aerial display quality and higher gesture interaction quality;
C) conventional display and higher gesture interaction quality.

The details of these features are as follows:

Structure A) This structure was designed such that the images were located as close as possible to the user and the display contrast became greater, based on the following approach. In the case where the image was located very close to the user, some people were uncomfortable in a manner similar to stereoscopic 3D systems. In this study, we performed a prior examination using twenty participants to estimate the closest location limit where this discomfort was not induced in the users. The limit distance from the Fresnel lens to the image position was about 400 mm in case the image observation distance was about 500 mm. The image location in this system was arranged so as not to exceed this limit. An optical polarizing film was installed on the bottom surface of the housing aperture glass to increase the aerial display contrast (in detail, see Appendix). The IR sensor beams were spread along the aerial display plane, and the perpendicular distribution was present to the extent of several tenths of a millimeter. In structure A, the distribution center was located approximately 20 mm to the back of the aerial display plane. Then, the sensitivity of the IR sensor became lower than the optimized condition at the aerial display plane. The probability the gesture recognition system did not operate when the users touched the aerial display became relatively larger. Overall, this structure was designed as having a higher aerial display quality and moderate gesture interaction quality.

Structure B) In contrast to the structure A, structure
B was designed for lower aerial display quality and higher gesture interaction quality. The image plane was shifted 40 mm behind structure A. Then the IR sensor beam was distributed closer to the user in the perpendicular direction to the aerial display plane. Therefore, the trigger signal was generated and the gesture recognition operated slightly before the user’s fingers or hand reached the display plane. As the user’s fingers or hand passed through the distribution center plane of IR sensitivity, the probability the recognition system did not operate became relatively smaller than that of structure A. The polarizing film was also removed. Then the incident light intensity into the gesture camera became larger, but the contrast of the aerial image was lower (Fig. 2B).

Structure C) Instead of an aerial display, a conventional 2D display was utilized in structure C (Fig. 2C). This configuration has already been commercialized in peripherals of the consumer game machines. The setup in this case was arranged so that the objects look similar in size to the structure A or B. The distance between the display and the gesture plane was about 1 m. The gesture recognition system was identical to that used in structure B.

2.2 Software content

We prepared three types of software that could be used in the 2D aerial display user interaction setup (Fig. 3A to 3C). In all these systems, the objects made using 3D computer graphics (CG) were moved or changed by the user’s operations through gesture recognition. The details of this system are as follows:

1) Menu (Fig. 3A): The aim of this content was a usability test to investigate the user’s response to the fundamental gesture operation of the user interface in this system. This content was designed as a startup menu in this system. From this menu content, the user could transit to all other contents installed in this system. The operation included a search action of the content by the rotation of a hexagonal cylindrical menu object and the selection of the content by clapping hands. On each side face of the hexagonal cylindrical menu object, a thumbnail for each content was depicted. The rotation direction was only clockwise. By clapping the hands, the user could select the content in front of the menu object and the system then transits to the selected content. This gesture action was also allocated to back transition from each content to the main menu. By operating this object as the startup
menu, the user first learns how to operate the touchable objects.

2) Gold fish (Fig. 3B): The goal of this content was to evaluate how much the users were aware of depth during the gesture operations. In this content, a goldfish swims around on the aerial display plane. If the user places their finger around the goldfish, it swims around the user's fingers and gradually approaches them. When the swimming speed of the fish is very slow, the user needs to hold their motion for several seconds. During the waiting period, the user must become very sensitive to the relative position of the fingers to the goldfish in the image plane (xy plane) and along the optical axis direction (perpendicular to the display plane). The structures A, B, and C have the same gesture recognition systems. However, the depth sensing function was not equipped and the user’s hands or fingers were detected only on one gesture plane. In this system, however, the ball and the table and net are displayed in the same aerial plane. All the CG objects in this content were drawn with a perspective representation. The parabolic trajectory and motion of the ball were also represented using perspective in the aerial 2D plane. As in real table tennis, a player predicts the trajectory of the ball by observing its 3D motion. In the systems containing the structures A and B, although the hitting position of the ball was designed to be in the aerial display plane and to be almost identical to that of the real table tennis, the player needs to predict the trajectory from the pseudo-3D perspective trajectory in a 2D plane. One of the points of this content was how differences are reflected in the user’s evaluations. On the other hand, in structure C, which is the same configuration as commercially available game machines, the ball hitting position was also in the gesture plane and was far from the display. By comparing the results for structures A or B and C using this content, the advantage of the gesture interaction combined with the aerial display can be evaluated.

3) Table tennis (Fig. 3C): This content was a simple table tennis game. In contrast to the goldfish dynamic, this content was prepared to evaluate the user’s dynamic depth consciousness. When the player hits the ball using a racket, the ball is programmed to return again to the front of the player so that player can continue the game like real table tennis. This content contains a 3D scene, and the user’s gesture action should be three-dimensional. In this system, however, the ball and the table and net are displayed in the same aerial plane. All the CG objects in this content were drawn with a perspective representation. The parabolic trajectory and motion of the ball were also represented using perspective in the aerial 2D plane. As in real table tennis, a player predicts the trajectory of the ball by observing its 3D motion. In the systems containing the structures A and B, although the hitting position of the ball was designed to be in the aerial display plane and to be almost identical to that of the real table tennis, the player needs to predict the trajectory from the pseudo-3D perspective trajectory in a 2D plane. One of the points of this content was how differences are reflected in the user’s evaluations. On the other hand, in structure C, which is the same configuration as commercially available game machines, the ball hitting position was also in the gesture plane and was far from the display. By comparing the results for structures A or B and C using this content, the advantage of the gesture interaction combined with the aerial display can be evaluated.

3. Experimental Procedure

We have performed a usability test using the After Scenario Questionnaire (ASQ) method\(^7\). This method has been developed to measure scenario-based usability immediately after experiments wherein a scenario comprises tasks that are related to each other. The ASQ has been used in several user interaction evaluations, for example, a cross display interaction evaluation\(^8\), and a smartphone usability study\(^9\). As the purpose of this study was to evaluate the user’s impression of the proposed system, the data were acquired as soon as possible immediately after the participant was tested.
hence, the ASQ method is suitable for evaluating the proposed system. Based on this method, a participant answered a questionnaire comprising three questions inquiring about the degree of "satisfaction," "effectiveness," and "efficiency" just after the completion of each task of the scenarios. The Likert scale was used to obtain the results. In this experiment, the scale was anchored at point 1 as strongly agree and point 7 as strongly disagree.

Three common tasks were assigned to each participant. The task scenarios are shown in Table 1. These tasks were a combination of the three software contents and the three configurations of hardware setup. The questions were modified from the original ASQ because this proposed system intended to be applied for entertainment purposes. Seventeen volunteers (age: 20-50, 5 females and 12 males) were used as participants. All participants were visually healthy and could perceive the aerial images.

The questions were as follows:

**Questionnaire for Task 1**
1. Overall, I am satisfied with the natural use for completing the tasks in this scenario
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA
2. Overall, I am satisfied with the freshness or the novelty to complete the tasks in this scenario
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA
3. Overall, I am satisfied with the intuitive operation when completing the tasks
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA

**Questionnaire for tasks 2 and 3**
1. Overall, I am satisfied with the ease of completing the tasks in this scenario
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA
2. Overall, I am satisfied with the natural use to complete the tasks in this scenario
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA
3. Overall, I am satisfied with the support information (the sign shows the aerial display plane on which the image is constructed) when completing the tasks
   - Strongly agree
   - Strongly disagree
   - Not applicable
   1 2 3 4 5 6 7 NA

Regardless of the hardware structures, a real object sign was set on the table top as shown in the inset in Fig.1 to indicate the gesture plane which corresponded to the image plane in case structure A and B. This sign could support the participant to recognize the image or the gesture plane. In the question 3 in task 2 and 3, the participants were asked about the usefulness of this sign’s indication.

**4. Results**

Fig. 4 shows the experimental results of the usability tests. The vertical axis shows the scores of the participants’ evaluations. The horizontal axis shows the tasks, the intents of the questions, and the structures of the hardware. The dots in the graph show the averages and the error bars indicate the 95% confidence interval.

**4.1 Analysis of each task**

With respect to the Shapiro-Wilk test, the results did not show a Gaussian distribution for all the questionnaire results. The Kruskal Wallis tests were performed for each question and each task to reject the...
null hypothesis and clarify the significance of three types of structures. The p-values were derived as shown in Table 2. The multiple comparisons using Steel-Dwass test was also performed. In Table 3, the results are shown only for the cases where the significance or the trend can be expected in the Kruskal Wallis tests. In the following section, we describe the results for each task of the software content.

**Task 1: Menu**
The aim of this task was to confirm the user’s response to the fundamental gesture operation of this system. From the data for the p-values based on the Kruskal Wallis test in Table 2, the results from the questionnaires considering the novelty and the intuitiveness of the approach had a significance that satisfied the general standard of \( p < 0.05 \). The p-values for the question dealing with spontaneity was \( 0.05 < p < 0.10 \), which may indicate the trend. On comparing structures A, B, and C for task 1 for all questions, structure B was determined to have the highest score. Through multiple comparisons, the structure B was determined to be significantly better than structure C in terms of novelty and intuitiveness. On the contrary, structure A was not as popular structure B in terms of the questions of spontaneity, intuitiveness, and novelty.

As mentioned in section 2.1, the aerial display plane in structure B was further from the user, but the gesture sensitivity was higher than structure A. Here, the higher gesture sensitivity meant shorter time lag and lower error rate gesture recognitions. Therefore, in the evaluation of spontaneity and intuitiveness, the gesture sensitivity played an important role. If the user’s gesture was recognized and its command was processed slightly before touching the aerial image, the aerial image user interface was considered better in terms of spontaneity and intuitiveness. Conversely, in the case of poor gesture sensitivity where touching a location further away from the aerial display is required, the effects of the aerial display on the spontaneity and intuitiveness were lower. Regarding the novelty, the aerial display user interaction that had a higher quality gesture interaction was also evaluated as being better. In the multiple comparison test results for the novelty shown in table 3, however the p-value between structure A and B was larger value than 0.10 and did not show the significance unlike for the spontaneity and the intuitiveness.

**Task 2: Goldfish**
The aim of this task was to evaluate how aware the users were of structure depth during the gesture operations. For all three questions in this task, the results of the Kruskal Wallis test did not reject the null hypothesis based on p value results. However, a few trends were observed by comparing the results with the Task 1. In terms of spontaneity in task 2, structures A and B were equal. But structure C seemed to be less favorable. In task 2, the user’s operation was to point the position on the aerial display plane in which case the user’s operability did not become important. In this task, the usefulness of the supporting information was also considered. The participants were asked about the degree of usefulness of the aerial image as an indication.

Through multiple comparisons, the structure B was determined to be significantly better than structure C in terms of novelty and intuitiveness. On the contrary, structure A was not as popular structure B in terms of the questions of spontaneity, intuitiveness, and novelty.

As mentioned in section 2.1, the aerial display plane in structure B was further from the user, but the gesture sensitivity was higher than structure A. Here, the higher gesture sensitivity meant shorter time lag and lower error rate gesture recognitions. Therefore, in the evaluation of spontaneity and intuitiveness, the gesture sensitivity played an important role. If the user’s gesture was recognized and its command was processed slightly before touching the aerial image, the aerial image user interface was considered better in terms of spontaneity and intuitiveness. Conversely, in the case of poor gesture sensitivity where touching a location further away from the aerial display is required, the effects of the aerial display on the spontaneity and intuitiveness were lower. Regarding the novelty, the aerial display user interaction that had a higher quality gesture interaction was also evaluated as being better. In the multiple comparison test results for the novelty shown in table 3, however the p-value between structure A and B was larger value than 0.10 and did not show the significance unlike for the spontaneity and the intuitiveness. Then, structure A seemed to be similar to structure B at the point of the novelty. The common point of structure A and B was the aerial display. Therefore, the visual effect was considered to play an important role in the novelty.

**Table 2** Usability test questionnaire results after analysis by Kruskal Wallis test. Values in red show rejection of null hypothesis.

<table>
<thead>
<tr>
<th>Contents</th>
<th>Question point</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menu</td>
<td>Spontaneity</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>Novelty</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>Intuitiveness</td>
<td>0.034</td>
</tr>
<tr>
<td>Goldfish</td>
<td>Ease of use</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>Spontaneity</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>Support</td>
<td>0.801</td>
</tr>
<tr>
<td></td>
<td>information</td>
<td>0.468</td>
</tr>
<tr>
<td>Table tennis</td>
<td>Ease of use</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>Spontaneity</td>
<td>0.428</td>
</tr>
<tr>
<td></td>
<td>Support</td>
<td>0.468</td>
</tr>
</tbody>
</table>

**Table 3** Usability test questionnaire results. The table shows multiple comparison results using Steel-Dwass test; values in red shows rejection of null hypothesis.

<table>
<thead>
<tr>
<th>contents</th>
<th>question</th>
<th>structures</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>menu</td>
<td>Spontaneity</td>
<td>A:B 0.10</td>
<td>A:C 1.00</td>
</tr>
<tr>
<td></td>
<td>A:C 1.00</td>
<td>B:C 0.15</td>
<td></td>
</tr>
<tr>
<td>novelty</td>
<td>A:B 0.44</td>
<td>A:C 0.19</td>
<td>B:C 0.03</td>
</tr>
<tr>
<td>Intuitiveness</td>
<td>A:B 0.07</td>
<td>A:C 0.98</td>
<td>B:C 0.05</td>
</tr>
</tbody>
</table>

Through multiple comparisons, the structure B was determined to be significantly better than structure C in terms of novelty and intuitiveness. On the contrary, structure A was not as popular structure B in terms of the questions of spontaneity, intuitiveness, and novelty.

As mentioned in section 2.1, the aerial display plane in structure B was further from the user, but the gesture sensitivity was higher than structure A. Here, the higher gesture sensitivity meant shorter time lag and lower error rate gesture recognitions. Therefore, in the evaluation of spontaneity and intuitiveness, the gesture sensitivity played an important role. If the user’s gesture was recognized and its command was processed slightly before touching the aerial image, the aerial image user interface was considered better in terms of spontaneity and intuitiveness. Conversely, in the case of poor gesture sensitivity where touching a location further away from the aerial display is required, the effects of the aerial display on the spontaneity and intuitiveness were lower. Regarding the novelty, the aerial display user interaction that had a higher quality gesture interaction was also evaluated as being better. In the multiple comparison test results for the novelty shown in table 3, however the p-value between structure A and B was larger value than 0.10 and did not show the significance unlike for the spontaneity and the intuitiveness. Then, structure A seemed to be similar to structure B at the point of the novelty. The common point of structure A and B was the aerial display. Therefore, the visual effect was considered to play an important role in the novelty.
of the gesture plane in terms of depth. The results and evaluation scores were relatively low and did not show any significant variation in the various structures.

**Task 3: Table tennis**

In contrast to task 2 using the goldfish, this test was performed to evaluate the user’s dynamic depth consciousness. This was the same as the goldfish test, so its significance could not be confirmed for all three questions using the Kruskal Wallis tests. As this software content was very familiar considering that it was based on a real sport, the ease of use seems to be better for all three structures. In this task, the player’s gesture action was dynamic so that they may not mind small differences in the depth of the aerial display and the gesture plane. In terms of spontaneity, the difference between structures A or B and C was smaller than that in the case of goldfish task. We supposed that this was because the participants had experience with commercialized games similar to structure C. Another possibility is that the participants may not have minded the large difference in the display and gesture positions because they concentrated on the amusement of this content. With regard to for the usefulness of the support information, the differences between structures A or B and C could not be confirmed in the usability test results. One of the reasons for this may have been player’s dynamic gesture action along the depth direction. The probability that the player’s hand passed through the gesture plane increased during the gesture action in this content task. Therefore, the player may not need the supporting information for the gesture plane.

Summarizing the comparison of structures in each task of software content, we can conclude the following:

1. Higher quality gesture interaction played an important role in the usability of the combined system with the aerial display.
2. The aerial display induced significant novelty when the users operated the aerial image using the task menu.

**4.2 Factor analysis**

To extract the dominant factor that influenced the user experience, we performed an exploratory factor analysis (EFA). The dominant factors can be derived by an EFA. Before applying the factor analysis, we deleted some questionnaire results from the original data set comprise nine items. The question dealing with ease of use can be assumed to measure the performance of the gesture interaction and not the psychometric factor. The question dealing with the support information was considered to be an indirect measure of usability. Next, we removed all four questionnaire results employing these two questions for structures A to C and for the tasks 2 and 3. The remaining five items were used for the factor analysis. Otherwise, in the case where all items were used for the EFA, clear dominant factors were not obtained.

**4.2.1 Exploratory factor analysis**

The covariance matrices of the participant’s answer distributions for all the setups from A to C were calculated. The eigenvalues and eigenvectors were derived based on principal components analysis. The number of the eigenvalues whose values were larger than those at the inflection point were used for the number of factors for the factor analysis. In this analysis, two factor models were selected as the basis. These two factors represented 73% of the total factor contributions. Numerical calculations were performed using the R language software function\(^{14}\). A maximum likelihood method was used for the optimization. As correlations were observed between the answer distributions, a promax rotation, which is an oblique rotation, was applied for factor loading. The analysis was applied to the answers of the 15 questions from each participant dealing with the spontaneity of all the tasks of contents and the novelty and the intuitiveness for each task in the menu for each structure. An analysis was performed on all questions from all participants without distinguishing the structures. Tables 4 and 5 show the factor loadings and the factor contributions resulting from this analysis.

<table>
<thead>
<tr>
<th>Contents</th>
<th>Question</th>
<th>Factor1</th>
<th>Factor2</th>
<th>communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>table tennis</td>
<td>spontaneity</td>
<td>0.917</td>
<td>-0.208</td>
<td>0.608</td>
</tr>
<tr>
<td>gold fish</td>
<td>spontaneity</td>
<td>0.669</td>
<td>0.026</td>
<td>0.473</td>
</tr>
<tr>
<td>menu</td>
<td>novelty</td>
<td>0.517</td>
<td>0.13</td>
<td>0.382</td>
</tr>
<tr>
<td>menu</td>
<td>spontaneity</td>
<td>-0.116</td>
<td>1.078</td>
<td>0.995</td>
</tr>
<tr>
<td>menu</td>
<td>intuitiveness</td>
<td>0.399</td>
<td>0.447</td>
<td>0.616</td>
</tr>
</tbody>
</table>

Table 4 Results of the exploratory factor analysis; Factor loadings are shown. Color of the number shows the range of the loadings.

Table 5 Results of the exploratory factor analysis. Factor contributions are shown for each factor.

<table>
<thead>
<tr>
<th>Factor contribution (SS loadings)</th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.727</td>
<td>1.422</td>
</tr>
</tbody>
</table>
By noting the large value of loadings in the tables, each factor can be characterized as follows:

Factor 1: In the case of the questions regarding the spontaneity of the table tennis and the gold fish tasks, these factor loadings exhibited larger values. These content of these tasks implied amusement or impression rather than functionality or operability. The factor loading for the novelty of the menu tasks was also large. This factor was considered as being attractive.

Factor 2: The factor loadings for the spontaneity and the intuitiveness of the menu content were large. Most of the tasks associated with the menu content were related to usability. The spontaneity and the intuitiveness were considered to be the important characteristics that were required for system operability. Consequently, this factor should reflect the degree of system operability.

The comparison of the structures A to C in the EFA are shown in Fig. 5. The horizontal axis and vertical axis denote the scores of factors 1 and 2 i.e., the attractiveness and the operability, respectively. Each system score was derived by averaging all of the participants evaluated scores for each system. In this graph, structure A and B exhibited higher values in attractiveness and the structure B demonstrated a higher value in operability. Taking each system structure into account, it was assumed that the direct interaction with the aerial image provided a novel, natural, and attractive user interface, whereas the appropriate gesture sensitivity was crucial for good operability.

### 5. Conclusion

In this study, a usability test was performed on a display system that combined an aerial display with gesture interaction using participants. By applying Kruskal Wallis test to the experimental results, it is found that higher quality gesture interaction played an important role in the usability. And the psychometric novelty of the aerial display was also revealed in the case where the users operated the aerial image from a software menu in the system. The analysis results also implied that the users were not aware of the depth of the detail in this system. The factor analysis results showed that the user interaction with the aerial image provided an attractive user interface, and that the appropriate gesture sensitivity was crucial for good operability.

We believe this type of user interaction utilizing the aerial display is a novel, natural, and intuitive human computer interaction system for future development.

### Acknowledgments

This study was supported by Grants-in-Aid for Scientific Research Japan. We acknowledge Mr. Kunio Fukuda of Hybrid System Corp. for the technical advice. We acknowledge all participants in Yamaguchi Laboratory in Tokyo Institute of Technology and their friends and families collaborate the usability tests.

### References

2) S. Maekawa, K. Nitta and O. Matoba, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6803, 68030B
3) Y. Makino, Y. Furuyama, S. Inoue and H. Shinoda, HapticClone (Haptic-Optical Clone) for Mutual Tele-Environment by Real-time 3D Image Transfer with Midair Force Feedback, #chi4good, CHI 2016, San Jose, CA, USA
8) J. Grubert, M. Heinisch, A. Quigley, D. Schmalstieg, MultiFi: Multi-Fidelity Interaction with Displays On and Around the Body: CHI 2015, 3933-3942
11) S.S. SHAPIRO M.B. WILK, An analysis of variance test for normality (complete samples)Åı, Biometrika (1965) 52 (3-4): 591-611. DOI: https://doi.org/10.1093/biomet/52.3-4.591, Published: 01 Dec. 1965
Appendix: contrast enhancement system for the aerial display

The optical configuration of the enhanced contrast system is shown in Fig. A. The direction of polarization of the polarization film was arranged parallel to the LCD light emission. The linearly polarized emission light \( I_0 \) from the LCD was transmitted through the film with the transmission ratio \( T_p \) to the parallel polarization of the film. Conversely, the perpendicular component of the randomly polarized ambient light \( I_1 \) was absorbed. The remaining transmitted parallel polarized light reached the internal surface of the housing and was absorbed or scattered. The scattered light polarization was randomly rotated and only the parallel polarization component was transmitted through the film. Denoting the optical absorption as \( \text{Abs} \), the scattered ambient light intensity can be expressed by Equations 1) and 2) for the case of the polarizing film installed as \( (I_{sp}) \) and not installed as \( (I_{sn}) \). In contrast, the source display light intensity decreased when the polarizing film was used. The source intensities with and without the polarizing film, \( I_{dp} \) and \( I_{dn} \) are given in Eqs (3) and (4). Therefore, the contrast with and without the polarizing film, \( C_p \) and \( C_n \) are given by Eqs. 5) and 6). In the case where the polarizing film was installed the aerial display contrast became \( 4/T_p \) times larger. Generally, \( T_p \) was approximately 50% in this system. Then, eight times higher contrast can be expected by installing the polarization film.

\[
\begin{align*}
I_{sp} &= 1/4 \cdot I_1 \cdot T_p^2 \cdot (1 - \text{Abs}) \\
I_{sn} &= I_1 \cdot (1 - \text{Abs}) \\
I_{dp} &= I_0 \cdot T_p \\
I_{dn} &= I_0 \\
C_p &= 4 \cdot I_0 / I_1 \cdot T_p^2 \cdot (1 - \text{Abs}) \\
C_n &= I_0 / I_1 \cdot (1 - \text{Abs})
\end{align*}
\]