Graphic Interface of Virtual Assistant with Adaptable Screen Projection

Auejin Ham
Daegu Gyeongbuk Institute of Science and Technology
Daegu, Korea
ham@dgist.ac.kr

Hyunjin Kim
Korea Advanced Institute of Science and Technology
Daejeon, Korea
hjkim17@kaist.ac.kr

ABSTRACT
Virtual Assistants are commonly used for various real-life situations such as turning on music. Majority of the Virtual Assistants communicates with users based on audio interface due to convenience of remote control. However, audio interface has its own limitation due to the sequential propagation of communicable information so that users cannot receive multiple information efficiently. For this reason, visual displays can be applied to virtual assistant to show variety of information in a graphical user interface at the same time. However, fixed visual interface has a problem of locational dependency, so that it is hard to receive information from the visual display in a distance while the audio interface of the virtual assistant is able to be used. To solve this problem, we suggest a new visual interface with projector rotation and human location detection. This new interface directs the screen near the user who make requests, and make the display easy to be seen.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces - Graphical user interfaces, Interaction styles

Author Keywords
Virtual Assistant, Projector Visual Interface, Projector Rotation, Kinect Room Detection

INTRODUCTION
As natural language processing and deep learning technology develops, virtual assistant has started to appear on the market, and they takes human language and do some real-life tasks. As they leverage natural language processing technology, user interface for virtual assistants are mostly based on audio interfaces. Users make requests to the virtual assistant by saying some requests with natural language, and the virtual assistant which receives the requests processes corresponding actions and take back the result by speaking natural language to the user.

While it is easy to understand directly when user listens the response of the virtual assistant by natural language, there is a problem on the audio-based request/response mechanism. The problem is that the information which virtual assistants gives is received only sequentially because of the basic property of the sound. For instance, if there is a service that reply to the user requests asking groups of restaurants that can deliver food, then the user need long time to get full list of restaurant information by sound. To mitigate this problem, some virtual assistants, such as Amazon Echo Show, shows their own visual interfaces on television or on the display attached to themselves showing various information simultaneously.

Visual interface with display monitor has significant benefits as described above. But it arise another problem. The visual image or texts are more dependent on distance or location than sound. If some obstacles are blocking in the middle of the user and the virtual assistant, users are hard to see the results shown on the virtual assistant's display, while they can listen what the virtual assistant says. While if the virtual assistant is far from the user, user can simply raise the volume of the audio interface on virtual assistant should to hear more clearly. Therefore, to deal with the same situation with obstacles, the visual interface on virtual assistant should increase size of the screen, which is harder to achieve than increasing sound volume.

In this paper, we introduce the background information about the idea we used to make prototype in section 2. After that, We explain the prototype design and physical and software settings for the experiment in section 3. experiment details and results are shown on section 4 and 5. further discussion and limitations on our project is explained in section 6. At the last section, we concludes with the summary of this project.

Figure 1. Amazon Echo Show with attached visual display [1]

ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

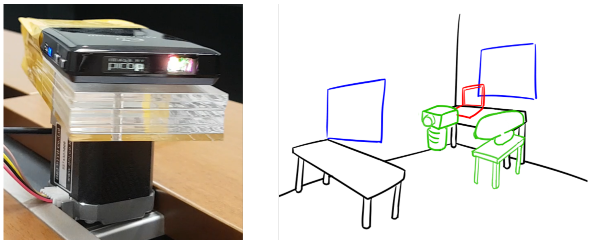

Figure 2. Prototype Design and Room Setting.

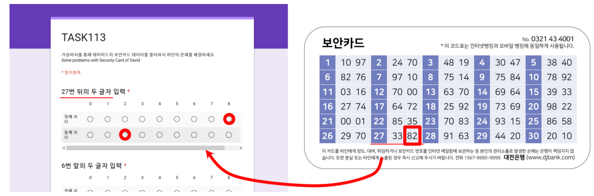

Figure 3. Task example of security card question.

BACKGROUND

To give new idea on the visual interface for virtual assistants, we first show some papers and works that is used to get an idea to set our prototype.

Visual Interface with Projector

Claudio Pinhanez [4] suggests the usage of projector to make alternative of carrying computer and setting up projector in the fixed screen positions. It uses mirror rotation to project display on the target position, so people project display where they want. Also, the paper consider about the issue of the brightness, project distortion, and focus problem on various projection area.

Room Environment Scanning

There are two works, RoomAlive [3] and IllumiRoom [2], which detects the room structure for interface usage. RoomAlive detects room environment and analyze room size, objects on the room with multiple depth camera. The analyzed room structure is used as a interface for display and input of augmented reality entertainment. IllumiRoom uses Kinect sensor to check the room structure, and it extends display screen to the surrounding of the screen without distortion caused by the objects.

PROTOTYPE DESIGN

To mitigate the locational dependency problem of fixed monitor display described in the section 1, we designed rotating projector display with Kinect room detection motivated by some aspects of the works in section 2.

Physical Prototype Design

For the physical design of our prototype, we use SST43D3000 step motor and SD02C motor driver. the motor is powered with 18V battery which is made by linear connection of two 9V battery. Since SST43D3000's holding torque is 3.55kgf-cm, and SD02C support maximum voltage of 20V, we need to select small projector to rotate without malfunction. We use MicroVision SHOWWX projector, and the projector requires battery and HDMI cable to connect with computer. The projector is attached on the motor like figure 2.

For rotation control, we use connect Arduino Uno to the SD02C motor driver with connecting to the serial port on a laptop. Arduino receives message from the laptop, and sending signal for enabling motor shaft, rotating motor, and setting direction of the rotation.

We install Kinect on the middle of the room to detect one side of the room. We only use one side of the room for doing experiment, we do not make Kinect to rotate to scan other sides of the room. The distance between Kinect and one side of the room is set under 4m to cope with the maximal distance of depth filter integrated in Kinect v1.

For comparison, we use 13.3-inch full HD display attached to LG PC gram 13 laptop as a counterpart of projector display.

Software Design

We implement control application with C#, which use Kinect depth filter display code. there's Kinect depth filter display shown on the control PC, and when we click a point to turn motor to make projector display screen on the point, it sends rotation signal with direction to the motor driver.

The program also calculates how many ticks the motor should rotate for showing display to the correct position. For this calculation, we assume that we already know the distance between Kinect and the motor is known. We also can get angle and distance information to the clicked point via Kinect. When we consider Kinect, motor, and clicked point is three points of triangle, since we know two triangle side's length and the angle between them, we can know other two angles. So the angle for motor rotation can be finally calculated from the Kinect data.

EXPERIMENT DESIGN

For experiment setup, we install one desk on the right side of the room and the other desk on the left side of the room. We set fixed laptop display on the desk on the right side. Overall experiment environment room setup is described on figure 2. Participants are using two desk on the figure as a experiment area, and participants cannot move to another desk while doing each task. The distance between the desk is around 3 meter.

Instead of using actual virtual assistant, we used Wizard of OZ method to do this experiment by using a presentation slides with TTS sounds and pretending actual virtual assistant is doing given tasks.

Total 4 people of participants took the experiments. Three participants are in the age group from 20 to 29, only one participant is in the age group from 10 19. All of the participants are students of the same university, and only one of the participant has an experience with virtual assistant.

Each participant do tasks within different types of devices and location of the participant, which means a participant does tasks on a desk on the right side and a desk on the left side with a laptop display. After that, the participant does the same tasks on each desks, with our projector display. The order of 4 trials are switched per two person.

We made security card writing task for experiment. We create

	Monitor		Projector	
	Near	Far	Near	Far
Avg (sec)	30.8	38.7	30.6	30.7
Std. dev (sec)	3.46	4.12	5.48	5.79

Table 1. average task completion time and standard deviation for each display interfaces and position of the user

Src	SS	DF	MS	F	P	F-test
Treat	368.1	1	368.1	23.25	8.11E-05	4.3
Err	348.3	22	15.83			
Total	716.5	23				

Table 2. ANOVA table of monitor display with independent variable of position of user, dependent variable of task completion time

Src	SS	DF	MS	F	P	F-test
Treat	0.042	1	0.042	0.0012	0.97	4.3
Err	763.6	22	34.71			
Total	763.6	23				

Table 3. ANOVA table of projector display with independent variable of position of user, dependent variable of task completion time

three sample securify card form, which has random number in each index. Each participant is asked to fill google questionaire form, like the figure 3. The form contains five questions that asks values in specific index of a security card. The participants should request virtual assistant to show specific security card data, and virtual assistant accepts the request and shows corresponding data.

We set total 12 trials of tasks, for two display devices, two position of the participants, and three individual security cards for doing task. For each trial, we measure task completion time. It is checked with recording each participant's request starting time, and google questionaire submission time

After the participants finish the task, each participant is asked to write TLX form. The TLX forms asks age group, experiance on virtual assistant, experiance on visual interface attached on virtual assistant, and how physically and mentally demanding the tasks are, for each task location and each visual display interface usage.

EXPERIMENT RESULT

There are the average task completion times for each devices and each location shown in table 1. The result shows that when the participants do task on the desk far from the fixed laptop display, it takes average 8 seconds more to finish the task than other situations.

physical demand rate	fixed monitor	projector
near mointor	2	2.75
far from monitor	7.5	2.75

mental demand rate	fixed monitor	projector
near mointor	2.5	2.75
far from monitor	4	2.75

Table 4. physical and mental demand rates for each category from 1 to 10 point

The standard deviation of the task completion time is higher when the participants use the projector display. We consider that this is because of the blurring projector display, which is caused by weak projector beam power and no automatic focus control.

We analyze the task time data with one-way ANOVA analysis. We separate each interface usage case and make two one-way ANOVA table with setting location as an independent factor. The table 2 and 3 show that when we use fixed monitor display, the p-value is small enough to say there is significant difference on task time when the location changes from near-display location to far-display location. On the other hand, There is high p-value for the projector display, so the position of the user does not have effect or have little effect on the task completion time for projector display.

We also summarize the TLX submitted from the participants. As shown on the table 4, the physical demands and mental demands are pointed around 2 $\tilde{3}$, except the task case on far-display location with fixed monitor display. For the case, the participants' physical demand point is average 7.5 point and mental demand point is 4 point. We think the result is because it is hard to see the display far from the user position. Also, participants should turn heads or body to the monitor display location to get information in the case, which makes more body movement than other task cases.

LIMITATION

Since we use small motor and motor driver that limits max voltage in 20V, because of the budget. The motor could not generate big force to rotate. So, we could not select big projector to use. Since small projector cannot display clear on bright room, we turn off light on the room when we use our prototype interface. Also, this limits the room size, so we cannot select ideal situation to show the location dependency of monitor display clearer than our testing room. So, for the future experiment, we can design better environment setting when we use a motor with higher power.

Also, HDMI cable and battery cable connected to the projector makes tension on motor rotation. The tension makes motor rotation inaccurate. To solve this problem, we can think about using projector that contains battery, and supports remote connection to the computer.

Finally, only four participants joined for doing experiment. To make more statistically reliable data, more participants should participate to generate reliable data.

CONCLUSION

We explained about the virtual assistant's sound interface and visual interface, and showed that there is locational dependency problem on visual interface. To mitigate the problem, we suggest new visual interface for virtual assistant, which uses projector display with motor rotation. We tested how the task time changes when fixed monitor display interface is changed to the interface that we proposed, and shows that the task time of our proposal is less dependent on location than fixed monitor display.

ACKNOWLEDGMENTS

We thank all the participants of the experiment, and all people who gave feedback on prototyping our idea on HCI class.

REFERENCES

1. 2017. Introducing the all-new Echo Show - Amazon Developer. (2017). https://developer.amazon.com/echo-show.
2. B. Jones et al. 2013. IllumiRoom: Peripheral Projected Illusions for Interactive Experiences. *Conference on Human Factors in Computing Systems* (2013).
3. B. Jones et al. 2014. RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units. *Annual ACM symposium on User interface software and technology* (2014).
4. Claudio Pinhanez. 2001. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. *International Conference on Ubiquitous Computing* (2001).