Volume 6, Issue 7, July 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Analysis Eye Gaze and 3D Hand Motion in Virtual Assembly Experiments

Nur Ani
Information System Of
Universitas Mercu Buana,
Jakarta, Indonesia

Sischo Fungky Aprilyan
Information System Of
Universitas Mercu Buana,
Jakarta, Indonesia

Cheng Cheng
Faculty of Computer Science
Beijing Institute of Technology,
China

Abstract– This paper reviews one of the advanced technologies in the field of human-computer interaction, the technology that is known as multimodal input. At the first we have to know the multimodal interaction. Many researchers in the field of HCI promise that one day the interaction between human and computer will be more natural, less error and more enjoyable. And whether this technology can give a maximal contribution in accordance its expectations itself? Specifically, the multimodal system can offer flexible, efficient, and natural interaction using input modalities such as speech, handwriting, hand gesture, and eye gaze. In this paper, we only focus on eye gaze and 3D hand motion. In this test, we use the virtual assembly system application that supports multimodal input of eye gaze tracking and 3D hand motion in order to test whether the combination of multimodal input will provide the desired effect or not. In this paper, we present the result from usability evaluation experiments on how the eye tracker and the 3D mouse are used by the users. And finally our result show that the eye tracker and the 3D mouse can be naturally used by the users.

Keywords– Human Computer Interaction, Multimodal Input, 3D hand motion, virtual assembly

I. INTRODUCTION

Multimodal interaction has an ultimate goal to make the interaction between human and computer more natural through the five human senses. Elaboration of further system multimodal is a system that offers a flexible, efficient and usable environment that allows the user to interact with the computer through input modalities, such as speech, handwriting, hand gesture and eye gaze. Hence, in this paper will tell about the multimodal input which is part of the multimodal interface, an input system needed in the field of virtual reality, etc. The main problem that will be discussed in this paper is the research of using multimodal input devices such as eye gaze tracker and 3D mouse can be naturally adopted by novice (non-experienced) users or not. The results obtained from this experiment are expected to be useful as a reference for other researchers in conducting research or development in the field of multimodal input related to implementation of multimodal input on a virtual assembly system, or could be wider in multimodal research inputs used in fields such as virtual reality gaming and to identifying opportunities such as multimodal input on a virtual assembly system to create the effect of naturalness and easy learn level.

For effective experiment, we set the scope of the problems only related with these activities: The experiment was tested by one or more users in the experiment of 3D mouse and eye gaze tracker; When eye tracking test that aided by The Eye Tribe hardware, and aided by software built that is useful to calibrate eye gaze for more accurate result; Then for 3D mouse I use **3D Connexion** device; Virtual assembly system used in testing using counting the time manually, with the aim that the results obtained are completely inconsistent with the truth; In this study also used the questions that will be addressed to the respondent to reinforce the results of research output; The first test was done by tester to picking up counting, matching feature and face matching using a 3D mouse; The second test we used picking up, feature matching and face matching computation in the multimodal input devices; Result of the final output is the verification of respondent's test result, whether combined multimodal input devices can provide natural result in usage as well as easy to use or not.

II. RELATED WORK

Research comprises "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of humans, culture and society, and the use of this stock of knowledge to devise new applications [3]." It is used to establish or confirm facts, reaffirm the results of previous work, solve new or existing problems, support theorems, or develop new theories. A research can also be an extension of field work in the past for the development of a work. To test the validity of instruments, procedures, or experiments, research may relate to elements of the prior project or the project as a whole. The main purpose of basic research is documentation, discovery, interpretation, or the research and development (R & D) of methods and systems for the advancement of human knowledge.

The main aim of the implementation is for supporting multimodal interface more flexible, powerfully expressive, and low cognitive load means of human-computer interaction. Some of the advantages that can be drawn from the application of multimodal interfaces obtained from are:

a. Robustness: Redundancy in multimodal input increases the quality of communication between the user and system, because conveying similar or related information through different modalities increases the likelihood of recognition [4]. Modalities work together to achieve a greater level of expressiveness by defining imprecision or modifying the meaning conveyed through another modality [4].
b. Naturalness: Multimodal communication can result in a high degree of naturalness, capitalising on the well-established practices of human-human interaction [4]. Complex tasks can be facilitated through use of multimodal interaction, because the paradigm effectively increases the communicative bandwidth between the user and system, increasing the level of input expressivity.
c. Flexibility: A major benefit of multimodal interfaces is flexibility, allowing for individual users to both perceive and structure their communication in diverse ways for specific context. Users can choose which modalities they employ, and how they are structured along semantic, temporal and syntactic dimensions.
d. Minimizing errors: Multimodal interfaces have been shown to increase performance by lowering the number of errors (through error avoidance), when compared to speech-only interfaces, in geographical interactive tasks [5]. Although multimodal interfaces have been shown to help users achieve shorter task completion times, particularly with map-based tasks, when compared to traditional GUI or speech-only interfaces, efficiency is not considered to be one of the major advantages of the multimodal paradigm. However, users can achieve faster error-correction when using multimodal interfaces [6].

The state-of-the-art multimodal input systems are currently only able to process two to three capital input channels are application-specific and have limited interaction vocabularies and grammars [7]. Many issues relating to synchronization, multimodal fusion (input interpretation), multimodal fission (output presentation) and interaction design is still in the discovery stage. Two input devices are already very mature in their development and are still in use today. Recently keyboard and mouse are being replaced by a blend of speech [7] and pen input speech and lip movement, or more [8]. Multimodal system that combines 3D user gesture modalities such as speech recognition recently come to the fore, however these are not mature when compared to technologies that use 2D recognizing input pen and ink modalities [9]. Naturally, the new challenges encompassed by gestural inputs including segmentation, feature extraction and interpretation of spontaneous movement trajectories, need to be addressed before such systems become more pravelent. The conclusion of the maturity of the multimodal system can we check from some multimodal system below, the system used multimodal pen / voice system developed by Cohen and colleagues [10] is the most mature of all four. Others include: VR Boeing Aircraft Maintenance Training System, developed by Duncan and colleagues [11]; the PEMMI system, a demonstrator interface developed by Chen and colleagues at NICTA [12], the which supports freehand using manual gestures together with speech, to monitor traffic; and the Portable Voice Assistant, developed by Bers and colleagues at BNN [13]. This information is needed in order to understand any system that allows the application of multimodal input.

III. DEVICES IN MULTIMODAL INPUT

There are two input devices that I use for research of eye tracking and 3D motion; these devices are Eye Gaze Tracker and 3D Mouse.

1. Eye Tracking System

Eye tracking is the process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head. An eye tracker is a device for measuring eye positions and eye movement. Eye trackers are used in research on the visual system, in psychology, in psycholinguistics, marketing, as an input device for human-computer interaction, and in product design. There are a number of methods for measuring eye movement. The first data regarding eye movements were obtained either through introspection or by the experimenter observing a subject's eye using a mirror, telescope or peep hole. These methods were dubious, of course, since any feature of the eye being studied could be obscured by the eye doing the studying. The first significant advance, therefore, was the invention of mechanical devices that would translate the eye's movements into permanent, objective records of its motion. The Eye Tribe Tracker is an eye tracking system that can calculate the location where a person is looking by means of information extracted from person's face and eyes. The eye gaze coordinates are calculated with respect to a screen the person is looking at, and are represented by a pair of (x, y) coordinates given on the screen coordinate system. A typical scenario is represented in Figure 1.

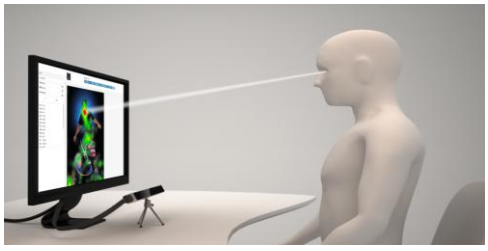

Fig. 1 User in front of monitor [21]

2. 3D Mouse System

The hardware used in this research of hand motion is a 3D Connexion mouse, a device like a mouse in general are like, but it has more full featured than the basic mouse in general. In the 3D mouse, there are additional buttons such as 1,2,3,4,5,6,7,8 , * , etc . Each button has a different function. This 3D mouse is devoted to the 3D model. In this mouse we can move the 3D models according to their needs.

3D Connexion actual motion controller is designed for conjunction with the mouse, allowing moves the 3D models to sync and screen on the screen. When we understand the movement of the models we will be ready to sync speed, control, and comfort work to be gained from using this 3D connexion mouse. And there are capabilities of *3D Connexion* mouse that is written in *3D Connexion* manual:

1. Move the model up and down
2. Moving the model left and right
3. Zooming the model in and out
4. Rotation the model on its Y axis
5. Tilt cap forwards/backwards to tumble the model on its X axis
6. Tilt cap left/right to rotate the model on its Z axis

Fig. 2 3D Mouse all Axis on Move or Zoom [22]

3. Virtual Reality

In a virtual environment in 3D space feels better than 2D, the human-machine interface is multimodal and the user will be entered into the computer-generated environment, the screen separating the user and the computer becomes invisible to the user. We are using Computer Aided Design (CAD) and computer aided engineering (CAE) software to validate a design before committing to making a physical prototype. Higher figures production have been developed in the form pro**to**types where engineering simulation software used to predict performance prior to constructing physical prototypes with the goal of engineers can quickly explore the performance of thousands design alternatives, without investing the time and money required to build physical prototypes. The ability to explore a wide range of design alternatives leads to improve the performance and design quality.

4. Virtual Assembly System

In the virtual assembly used in this study features that are used quite a lot, in addition to support 3D mouse devices, this system also apply assembly models using eye tracker. So the system is suitable for use in this study. The 3D models contained in the system database that will serve as the parts of a virtual assembly are:

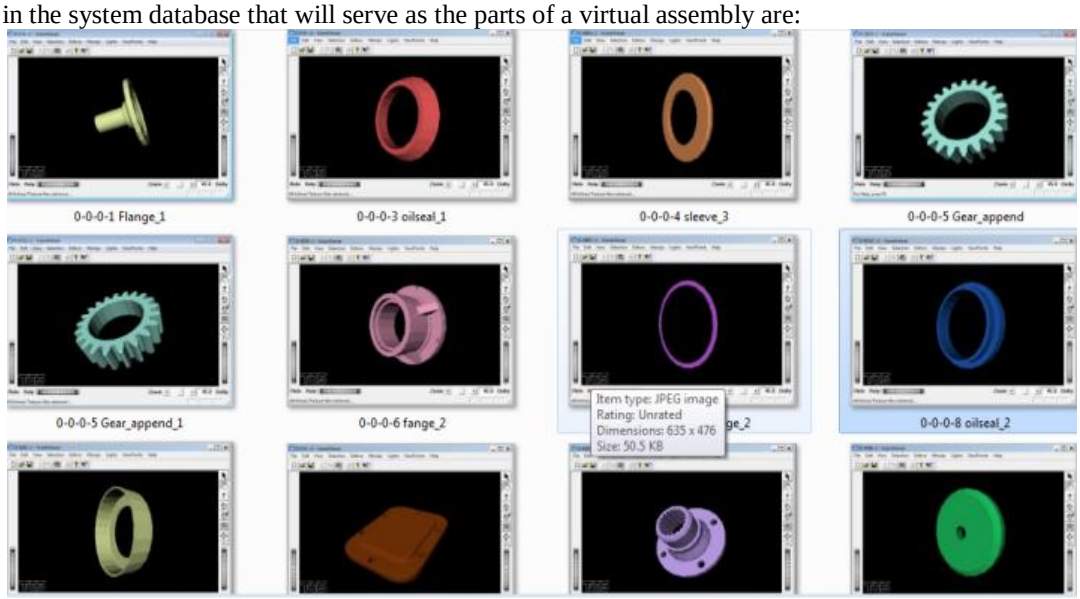

Fig. 3 Parts in Virtual Assembly System

5. Visual C++ and Open Inventor

The Open Inventor from Mercury is an object-oriented toolkit that eases the task of writing graphics programming into a set of easy to use objects. Then, since the Open Inventor is an object-oriented (written in C ++), It will encourage

programmers to expand the system by writing new objects. Open Inventor users have created new objects that previously did not exist, such as: Bezier surfaces, CSG objects, animation objects, special viewers, and many more. In the mechanism of Open Inventor rendering employs OpenGL objects. During the process of rendering each object to automatically create the proper efficient way to call OpenGL. Since Open Inventor was designed to uses OpenGL as graphics rendering, then is expected to take all the advantages of all OpenGL accelerators. Furthermore, Open Inventor rendering men support caching scheme roommates automatically generates very fast renderings of the database traversal without additional overhead. Open inventors including several interactive 3D objects and tools to create interactive objects, such as: Manipulators, draggers, and Projectors. The model of flexible event encourages programmers to extend the system to support new devices and event types.

IV. RESEARCH METHOD

In this study system the most important thing in this research is the ergonomic factor (human), Experiment Including picking up process, feature matching process, and face matching process in the virtual assembly system. The second factor is what affects whether the input is used to make the connection between human and virtual assembly system can produce a desired output or not.

The steps are performed in the data retrieval virtual assembly system as follows:

1. Respondents placed in front of a PC screen, where the Virtual Assembly System will appear on the PC screen.
2. An Eye Tracker System that is placed under the screen, where it is one of the most important stage that affects the accuracy of the calibration will affect the accuracy of eye gaze that will look at the monitor output. Therefore Eye Tracker System is placed in such a way to follow the needs of the user.
3. A 3D mouse also prepared near the user, placed in accordance with user demand.
4. Set up a camera to take photographs of the experiment.
5. The time given to the first process Picking Up is 10 seconds and the second process is 10 seconds, the first process in the face matching 10 seconds and the second process is 10 seconds, and the face matching 5 seconds. This is done to stimulate the respondent to use his cognitive side well.

In addition to the arrangement of the data collection stage through the Virtual Assembly System above, there are some important things that must be considered in order to get the desired data retrieval.

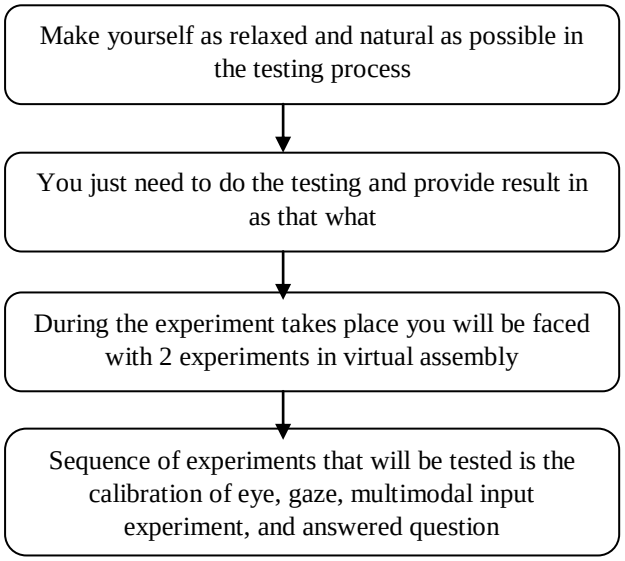

Fig. 4 Flow Attached Instructions

Research Process of Virtual Assembly System

Virtual Assembly System itself is a very important process in this experiment. Virtual assembly process itself has components which in turn can have a very important research data, while the components of the virtual assembly system in question are as follows:

1. Picking Up

Picking up is a process of moving a part from one position to another. This process is the initial process is performed in the assembly process. Just as the design is done on the physical assembly, this process is a process whereby a person must move the object to be assembled to the place of assembly operations either through manual method or using a production line system which was created to facilitate the process of virtual assembly.

In the experiment multimodal input for this paper, we need a method to determine the speed and accuracy of eye gaze and multimodal input through picking up hand motion. Picking up can be done by hand with the aid of 3D motion mouse or also can be done also with eye gaze and hand motion with the help of 3D mouse and also eye tracker which is two input devices that are used in this study. But in the present study focuses solely on observations of the performance

of multimodal input we can know by combining two elements, namely human senses of eye gaze and hand gesture to virtual assembly system. Do multimodal input is expected to be able to fulfill its duty in creating a multimodal input that has a good level of naturalness and also easy to learn for a beginner? So that users can moving objects with virtually the entire natural only through the help of eye gaze as the main parameter setting in motion as the mouse cursor with the help of 3D mouse as an auxiliary parameter in choosing the desired object by clicking on certain buttons.

In addition to using the time to determine the level of cognition load respondents and also to reinforce the results of data from respondent's answers to the questions given, there are several questions aimed to determine the level of naturalness and easy learn from both respondents.

To test the level of naturalness and also cognition load in this study the questions that will be included are:

a. Does the respondent can choose and moving parts from one position to another natural or not.
b. Did you ever know what a virtual assembly?
c. Do know what it is picking up in virtual assembly?
d. Is it easy in the operation of multimodal input device in this process?
e. Which part in this process which is considered the most difficult?

2. Feature Matching

Feature matching is a continue process of picking up, where objects that have been moved from one place to the destination is towards other objects related and can be combined experience the process of merging. Not all objects can be linked directly, these parts must be combined with a match part, and in this stage the user can find a match part that marked red. After finding a suitable partner, the next process is to determine the proper axis is marked with two red axis of the two objects have a position parallel with each other. A part in virtual assembly system it actually has three distinct colors axis, that red axis, blue axis and green axis.

Fig. 5 Red Axis in the Two Parts and Two others Axis in the Two Parts become Yellow

3. Face Matching

Face matching in Virtual Assembly System is a process that is done after a successful feature matching process. Face matching is the process of placing the main parts (parts that have to be combined with the part of interest) to the right position and predetermined by the designer, so we cannot position the main part in a position that was not designed by a designer. The process of face matching also requires 3 same axis with the feature matching, ie axis of red, blue and green. And the red axis in this process indicates the two parts have been matched and the main part will only be able to move in the direction of the red axis.

Then there are additional embedded features in the face matching that is like a square that contained in both parts. The main part can adjust the position of the auxiliary part, where it is based on the form of assembly what we want to create. The more detailed explanation we can see from the figures below:

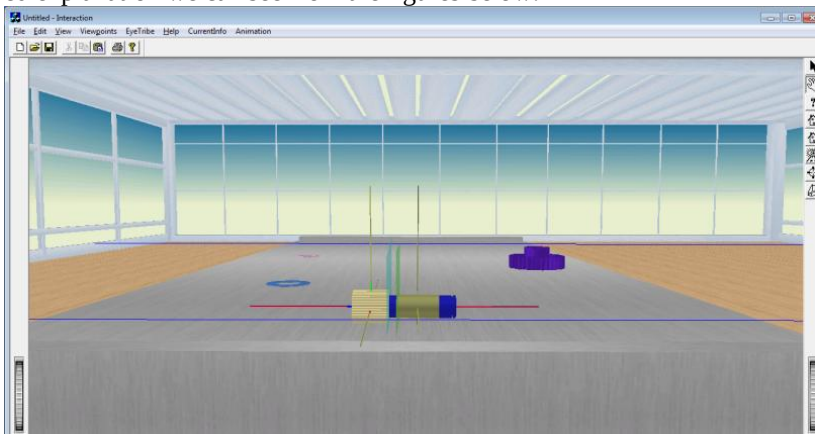

Fig. 6 Face Matching in the Edge Intent

V. DISCUSSION

User Determination

As already described in the previous chapter, there is an observation of the target of user that is useful to establish criteria of the respondents, the respondent in accordance with the requirements of the research will greatly influence the success of an observation, while the selected criteria to establish the proper respondent is the following:

a. Category user taken in this experiment is the category of non-experienced users who will test whether the multimodal input used on a virtual assembly system easy to learn or not.
b. Respondents consisted of Beijing Institute of Technology's Student in Computer Science department.
c. Doing measurement time during in this experiment, the time measured in order to consist whether the time spent testing process according to estimates determined by the researcher or not. This is to determine the level of cognition load the user in understanding this test, and also to support the response of users to the questions given.
d. At the end of the experiment will be given of questions aimed at obtaining data which is expected to support the research. Question is described in the previous chapter.

Multimodal Devices Testing

Before we can test the multimodal input devices on a virtual assembly system, there are stages that must be passed by the respondents include the calibration process at The Eye Tribe is very important before the other processes in the virtual assembly system. It is intended to introduce the respondent with the eye tracker device, of course, is very useful to give guidance on how to use the eye tracker device to a non-experienced user. Then in addition to eye tracker system that we want to introduce the course, there are other devices that 3D connexion. Perhaps many have not familiar with this device. For this reason in this section we will discuss about the impression of this device.

- Eye Tracker and 3D Mouse

In experiments conducted on 3D mouse by respondents who have never used a mouse before, there are several 3D impression gained from the first experience of the respondents. The main factors that affect the performance of the start of the respondents is that they do not know the function of each button located. And 3D mouse devices also sensitive level of high mouse cursor makes the respondents somewhat difficult to use 3D mouse at the start of use. The initial phase should be done by the respondent is in the eye calibration of The Eye Tribe user interface , the results of the calibration eye will affect the testing of multimodal input in the virtual assembly.

Table 1. Data of Respondents Calibration

No	Name	Category	Eye Calibration Result
1.	Respondent 1	Non-Experienced	5 Stars (Perfect)
2.	Respondent 2	Non-Experienced	4 Stars (Very Good)

Testing in Virtual Assembly

Experiments in this process require the respondent to work harder in comparison to the production process. Due to this process is the initial process for respondents in trials of multimodal input system to be tested on a virtual assembly system. There are two main processes in Picking Up process and while the details of the process of picking. The results of testing on the first process we can see the results in the table below:

Table 2 Picking up Timing Result for the First Process

No.	Respondent	Time
1.	Respondent 1	7.1 s
2.	Respondent 2	4.5 s

Feature Matching

In the process of feature matching, there are several steps that must be passed by the respondent.

a. There is a process of determining which parts of the second part that can be entered or occupied by the main part.
b. There is an axis adjustment process as described in chapter 3.

For that there are two steps in the process of feature matching experiments that tested by the respondent. First Step respondents were asked to specify which part of the part of interest that can disassemble by the main part. It is characterized by multiple sections or a part of the second part which turns into red.

Table 3 Feature matching timing result for the first process

No.	Respondent	Time
1.	Respondent 1	8.6
2.	Respondent 2	13.4

In the figure 7 is taken from the main part of the testing process respondents marked with a blue arrow can make the process of matching feature to the part of interest marked with red arrows. It can be seen from the section of the part of interest which some red to follow. The test results of the experiments in this step are described in the Table 3.

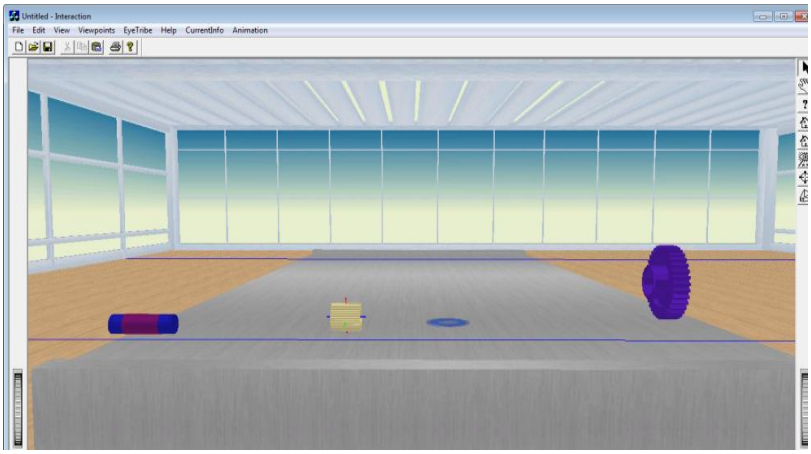

Fig.7 the First Feature Matching Process

Face Matching

Face matching is the final process in the virtual assembly of this system. Process to be followed respondent in accordance with chapter 3 is to match the position of the main part to the part of interest on the condition that has been successfully matching feature in the previous stage. The result of of the timing we can see from the table 4.

Table 4 Face matching timing result

No.	Respondent	Time
1.	Respondent 1	1.5 s
2.	Respondent 2	5 s

The results of the counting process time we can see in the table above where responden1 has a very fast time with a time difference that is far less than the respondent 2. In this test respondents can quickly understand and carry out instructions given by the examiner.

VI. CONCLUSION

In proving whether multimodal input system has a level of good naturalness or not so in the experiment result conducted by two of the respondents non-experienced, it was concluded that the multimodal input system is a breakthrough in which the system has a degree of naturalness that is quite good as evidenced by the answers of respondents, who argue and disagree that the fusion of these two devices in accordance with what is thought by the brain. When the eye gaze look at a certain part then quickly driven intent contained in the virtual assembly system can understand what the user wanted through eye gaze and hand motion. And the respondents who had not previously been tested were amazed by this experiment.

Then in proving whether cognition load required at this experiment in managing attained, or respondent with overload can use load less than the previous expectation of researcher who are experienced in using both these devices simultaneously (multimodal input). In this experiment whether multimodal input on a virtual assembly system easy to learn or not? And it turns out the experiment results collected in the data on each of the testing process can be concluded that the respondents could understand all instructions testers well, then test well were also supported by data computation time to prove the answers of all the questions to the user.

So overall mix of multimodal input devices by applying a mix of eye gaze and hand motion can giving a natural effect and is also not difficult to be learned and applied to the virtual assembly system is tested on the test this time. Virtual assembly system used can be said to be an excellent, proven system can read the movement of eye gaze through the help of The Eye Tribe tracker device and also the movement of hand motion that can be read properly by the virtual assembly system.

REFERENCES

[1] Regina Bernhaupt, Philippe Palanque, Marco Winkler and David Navarre. (n.d.). Usability Study of multi-modal interfaces using Eye-Tracking. IRIT, Toulouse III, , 1-14.

[2] Ferri, F., & Paolozzi, S. (2009). Analyzing Multimodal Interaction. In P. Grifoni (Ed.), Multimodal Human Computer Interaction and Pervasive Services (pp. 19-33). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-386-9.ch002

[3] OECD (2002) Frascati Manual: proposed standard practice for surveys on research and experimental development, 6th edition. Retrieved 27 May 2012

[4] Mark T. Maybury and Wolfgang Wahlster, editors. Readings in Intelligent User Interfaces. Morgan Kaufmann Publishers, 1998.

[5] Sharon Oviatt. Multimodal interfaces for dynamic interactive maps. In Proceedings of the ACM CHI 96 Human Factors in Computing Systems Conference, pages 95–102, Vancouver, Canada, 14-18 April 1996. ACM, ACM.

[6] Philip Cohen, David McGee, and Josh Clow. The efficiency of multimodal interaction for a map-based task. In Proceedings of the sixth conference on Applied natural language processing, pages 331–338, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[7] Sharon Oviatt and Philip Cohen. Perceptual user interfaces: multimodal interfaces that process what comes naturally. Commun. ACM, 43(3):45–53, 2000.

[8] D. G. Stork and M. E. Hennecke, editors. Speechreading by Humans and Machines. Springer Verlag, New York, 1995.

[9] L. Miguel Encarnacao and Lawrence J. Hettinger. Perceptual multimodal interfaces. IEEE Computer Graphics and Applications, 23(5), 2003.

[10] Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen, and Josh Clow. Quickset: multimodal interaction for simulation set-up and control. In Proceedings of the fifth conference on Applied natural language processing, pages 20–24, San Francisco, CA, USA, 1997. Morgan Kaufmann Publishers Inc.

[11] L. Duncan, W. Brown, C. Esposito, H. Holmback, and P. Xue. Enhancing virtual maintenance environments with speech understanding. Technical report, Boeing M& CT TechNet, 1999.

[12] Fang Chen, Eric Choi, Julien Epps, Serge Lichman, Natalie Ruiz, Yu Shi, Ronnie Taib, and Mike Wu. A study of manual gesture-based selection for the pemmi multimodal transport management interface. In Proceedings of the 7th International Conference on Multimodal Interfaces, ICMI 2005. ACM Press, October 2005.

[13] J. Bers, S. Miller, and J. Makhoul. Designing conversational interfaces with multimodal interaction. In DARPA Workshop on Broadcast News Understanding Systems, pages 319–321. DARPA, 1998.

[14] Mark T. Maybury and Wolfgang Wahlster, editors. Readings in Intelligent User Interfaces. Morgan Kaufmann Publishers, 1998.

[15] Sharon Oviatt. Ten myths of multimodal interaction. Communications of the ACM, 42(11):74–81, 1999.

[16] D. G. Stork and M. E. Hennecke, editors. Speechreading by Humans and Machines. Springer Verlag, New York, 1995.

[17] Sharon L. Oviatt. Multimodal interfaces. In J. Jacko and A. Sears, editors, Handbook of Human-Computer Interaction: Fundamentals, Evolving Technologies and Emerging Applications (revised 2nd sdition), pages 413–432. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2 edition, 2008.

[18] S. Bangalore and M. Johnston. Integrating multimodal language processing with speech recognition. In B. Yuan, T. Huang, and Tang X., editors, Proceedings of the International Conference on Spoken Language Processing (ICSLP2000). ACM.

[19] L. Wu, S. Oviatt, and P. Cohen. Multimodal integration - a statistical view. IEEE Transactions on Multimedia, 1(4):334–341, 1999.

[20] Jong, T. d. (2009). Cognitive load theory, educational research,. University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands. 105-134.

[21] Manual book of The Eye Tribe Tracker [Online], available: http://dev.theeyetribe.com/cpp/

[22] Manual book of 3D Connexion [online], available: http://www.3dconnexion.co.uk

