Journal of the Ergonomics Society of Korea
Vol.19, No.2 2000. 8. 31

Three-dimensional Teleoperation Performance Measures: Their correlations and effects of task difficulty*

Sung Ha Park**

ABSTRACT

Present study was performed to compare different types of teleoperation performance measures. A modified Cooper-Harper rating scale and a distance measure were newly introduced and compared to two traditional measures, the error rate and time-to-completion. Participants performed a simulated telerobotic task with varying visual display interfaces. The results of correlation analyses revealed that the modified Cooper-Harper ratings had a close association with the error rate measure. The distance measure also appeared to be a consistent performance measure, which accounted for both the error rate and time-to-completion. A further analysis, however, showed that the effectiveness of each performance measure varied with the level of task difficulty. The results imply that the choice of performance measure should be made based on the task difficulty as well as the particular task being studied. The results revealed here also have implication for performance measures of various display and control systems including aircraft navigation displays and tactical displays.

Keyword: Teleoperation, Performance Measure, Telerobotics, Displays

* 본 논문은 1999년 한남대학교 교비 학술연구비 지원으로 수행되었음.
** 한남대학교 산업공·기계공학부
주소 : 306-791 대전광역시 대덕구 오정동 133
전화 : 042-629-7535
E-mail : shpark@eve.hannam.ac.kr

1. Introduction

Tele-robots are devices with mechanical arms for manipulating objects within a working environment based on controlling inputs received from human operators (Sheridan, 1992). A typical tele-robotic task is the selection, manipulation, and placement of objects in a manufacturing area or storage facility (often called a pick-and-place task). Human exposure to hazardous tasks can be reduced when teleoperation systems are used to handle those tasks: the human operator is remotely located from the source of danger.

Teleoperation has been used in a wide variety of applications. Examples of such systems include applications in surgery, undersea pipeline installation, hazardous material handling, bomb disposal, and construction in space (Sheridan, 1992). Current teleoperation systems are becoming more complex and may be automated or semi-automated. However, the role of human operators will continue to be required for teleoperational tasks, which cannot be easily automated due to operational uncertainty or technological limitations. Moreover, the human operators must play an important role as backup when an automated system fails.

The necessity of human involvement and role in teleoperated systems has motivated the present study on the comparison of different teleoperation performance measures. Various performance measures have been used in the field of teleoperation. Two generally accepted measures are the accuracy (measured by counting the number of errors) and speed (measured by the time-to-completion). Many studies employed the accuracy as a teleoperation performance measure (Hannaford, Wood, McAffe, and Zak, 1991; Brouse and McDonalds, 1992; Kugath, 1972). There are also many experimental studies which consider the time-to-completion as a teleoperation performance measure (Kim, Tendick, and Stark, 1987; Kim and Stark, 1989; Massimino and Sheridan, 1994).

Fitts (1954) asserted that the time-to-completion had a positive linear relationship with task difficulty. Fitts' law has been used to describe human motor control performance. It is not clear, however, that Fitts' law can be used to describe the telemanipulation control performance. McGovern (1974) found that Fitts' index of difficulty was suited for describing human teleoperation performance. Draper, Handel, and Hood

(1990) also showed that teleoperation conforms to Fitts's law, in that movement time for a tapping task between two targets is highly correlated with the index of task difficulty. Conversely, its accuracy has been questioned in other experiments (MacKenzie, 1989; Massimino and Sheridan, 1994). Massimino and Sharidan (1994) found that the time-to-completion for a peg-in-hole telemanipulation had an exponential relationship with the task difficulty rather than a linear relationship.

With the controversy on its accuracy, The time-to-completion measure does not necessarily represent the quality of the teleoperation task being performed (McLean, Prescott, and Podhorodeski, 1994). Frequently, utmost safety consideration in critical applications (e.g., nuclear application) precludes the use of the time-to-completion as a teleoperation performance measure, favoring the use of accuracy measure.

Subjective measures, including rating scales, have been widely used as workload assessment techniques. Subjective ratings require human operators to rate their own experience of the workload after completing a given task. Major advantages of subjective measures are that they can be relatively nonintrusive and do not disrupt primary task performance if implemented correctly (Eggemeier, 1981). Few studies, however, have attempted to use the subjective measure as a teleoperation performance measure. This may be due to the lack of reliable data related to the use of subjective measures in teleoperated systems. In addition, some studies showed the disassociation of performance and subjective ratings. For example, Wickens and Yeh (1983) showed that subjective measures could be less sensitive than performance based measures derived from operator's behavior.

Another way of evaluating a teleoperation system is to take measurements on telemanipulator motion effort. These measures are based on the idea that human manipulation of a teleoperated system using a display and control interface should result in and be reflected to telemanipulator motion. In this regards, McLean, Prescott, and Podhorodeski (1994) suggested that the performance of a teleoperation system could be evaluated by analyzing manipulator motions. They performed an experiment by employing a peg insertion telerobotic task in a dynamically varying environment. The results showed that performance measures based on robot's joint effort and end-effector velocity were highly correlated with the time-to-

completion and subjective measures. They asserted that the manipulator motion based measures had a clear advantage of not relying on a task with a predefined end point, as compared to the time-to-completion measure.

The research presented in this paper was conducted to investigate and compare techniques of measuring teleoperation system performance. Traditional speed and accuracy measures were compared to a modified version of the Cooper-Harper Rating Scale (Casali and Wierwille, 1982, 1983; and Rahimi and Wierwille, 1982), and the distance that the end-effector traveled. The Modified Cooper-Harper Rating Scale is a 10-point scale ranging from 'very easy' at 1 to 'impossible' at 10. The subject is led through a series of structured questions to arrive at a final rating. End-effector distance was measured by calculating the distance of the moving end-effector from the beginning to the end of the task in three-dimensional space. These different performance measures were compared by employing a simulated telerobotic system with different visual interfaces. The effects of task difficulty on the effectiveness of performance measures were also of primary concern.

2. Method

2.1. Subjects

Five male and five female college subjects served as participants in this study. The subjects had 20/20 or corrected to normal visual acuity and full color vision. In order to minimize learning effect and individual difference, subjects were trained until their learning curves and performance met a certain criteria established based on the analysis of percent learning through a pilot study.

2.2. Apparatus

The equipment used for this experiment included a SiliconGraphics Indigo 2 (SGI Indigo 2) workstation with a 21″ monitor with a resolution of 1280×1024, a six-degree-of freedom Spaceball® 2003™(as a controller interface), and Crystal Eyes 2 eyewear with an emitter (as a stereoview interface). The experimental task environment was simulated using CAD, DEVICE, and WORKCELL contexts of TELEGRIP™ (release No. 3.0, Deneb Robotics, Inc) software and implemented using C language.

2.3. Simulation of the task environment

Visual displays for the telerobotic task included a force-torque display and a display for the work environment. The force-torque display provided information on the three-translational and three-rotational force levels applied by the Spaceball® 2003™. The displays for the work environment represented the current state of the robot (RX-130) and its working environment as human operators controlled the robot.

As a six-degree-of-freedom control input device, the Spaceball was used to manually control the robot and objects attached to the robot. The simulated robot was controlled by the Spaceball utilizing the end point control. To apply the end point control, an inverse kinematic algorithm built into the TELEGRIP was modified for the RX-130. As a stereoscopic view interface, the stereo eyewear, CrystalEyes 2 was installed to achieve the stereoscopic view. All of the experimental variables were manipulated and implemented by using the CLI (command line interpreter) commands in the C codes.

The experimental task was to control the robot (RX-130) arm for picking-up a virtual object (actually picking-up the tip of the object in the middle) on a table and placing it into a storage rack in the middle (see Figure 1).

2.4. Experimental Design

The experimental design was a repeated-measures design with three levels of display type, four levels of visual enhancement, and two levels of task difficulties as independent measures(within-subjects). Three display types included multiple 2-D, 3-D perspective (monocular), and 3-D stereoscopic display. Figure 1 shows the multiple 2-D display used in the experiment.

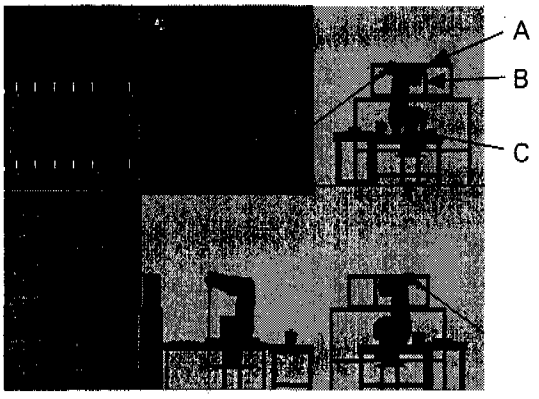

Figure 1. Multiple 2-D display with a solid reference line used in the experiment (A: storage rack, B: robot RX-130, and C: virtual object). From top-left, clockwise: Force-Torque Display, Plan-View, Right Side-View, Left Side-View, Front-View, and Task Status Display. Each view was 2-dimensional and orthogonal to each other.

In order to aid operators in perceiving depth, four visual enhancement cue types were applied to each display condition, including no visual enhancement cue, one solid reference line, one translucent reference cylinder with the solid reference line, and four reference lines with the solid reference line (Park and Woldstad ,1998).

Two levels of task difficulty were manipulated by changing the size of the cell of the storage rack. The combination resulted in 24 display configurations, which were differed in terms of depth information provided through the computer screen. The type of control was not considered as an experimental variable and fixed to an end point control using the Spaceball. The dependent measures included error rate (i.e., accuracy), time-to-completion (i.e., speed), distance (moved by the end-effector), and subjective ratings of perceived workload as discussed previously.

3. Data Analyses and Results

The experimental design resulted in 240 (3×4×2 display conditions and 10 replication) data points for each performance measure. The effectiveness of three visual display formats and four types of visual enhancement cues were not of primary concern in the present study. Briefly summarized, the results of repeated measures analysis of variance showed that a multiple two dimensional display was better than perspective and stereoscopic displays ($p<0.01$). For the 3-D display enhancements explored, the translucent cylinder enhance was clearly the best, with the line enhancements better than the no enhancement condition, but not as good as the TRC enhancement. If the perspective or stereoscopic displays are employed, visual enhancement cues should be used to assist with depth perception (Park and Woldstad, 2000).

The association of four different performance measures with varying visual display condition was evaluated in detail by investigating correlation coefficients. Table 1 shows correlation coefficients between four different performance measures compared in this study. The subjective based measure indicates a higher correlation with the baseline measure of error rate ($r=0.91$) than with the time-to-completion measure ($r=0.57$). During the experimental session, subjects were instructed to finish the task as fast as possible while minimizing the number of errors. Although the time-to-completion and error rate were given equal

Table 1. Correlation coefficients between performance measures.

	Error rate	Time to completion	Subjective ratings	Distance
Error rate	1			
Time to completion	0.76	1		
Subjective ratings	0.91	0.57	1	
Distance	0.90	0.87	0.80	1

importance, it seems that subjects relied more on the error rate (i.e., accuracy) than the time-to-completion in rating their perceived workload.

As an end-effector motion based performance measure, the distance measure is highly correlated with both the error rate and time to completion measures ($r = 0.9$ and 0.87, respectively). A possible explanation for the high correlation is that with errors operators need to perform additional corrective movements of the end-effector and therefore require more time. The plots in Figure 2 show examples of end-effector movement profiles with two of 12 display configurations under the high task difficulty condition (i.e., the multiple two dimensional display and the 3-D perspective display configurations without any visual enhancement cues). In this example, using the perspective display, the operator committed ten errors, rated perceived workload as '7 (Major difficulty)', and spent 303 seconds to complete the task. The distance moved by the end-effector was 10.54 meters. Using the multiple two dimensional display, however, the same operator committed one error, rated perceived workload as '2 (Easy, desirable)', and spent 173 seconds to complete the same task. The distance moved by the end-effector was 6.37 meters. Two plots in the Figure 2 show a similar trend of end-effector movement. However, as compared to Figure 2 (b), Figure 2 (a) indicates a longer decision time in changing direction of the end-effector and a large variability in adjusting the end-effector position (e.g., unnecessary up and down movements of the end-effector in y direction in three dimensional space).

Sensitivity, one of the criteria for the choice of workload assessment methods, means the capability of a method to discriminate changes in the workload

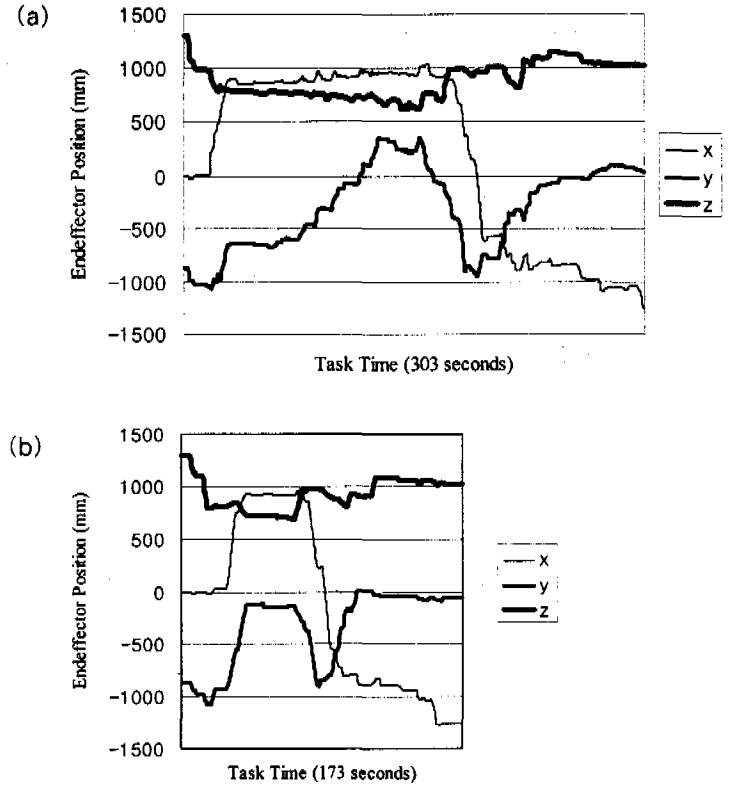

Figure 2. End-effector movement profiles in three dimensions (x, y, and z) under the high task difficulty condition. (a) Task with a 3-D perdpective display without any visual enhancement cues and (b) Task with a multiple 2-D display without any visual enhancement cues.

imposed by a task (Wickens, 1992). In general, the sensitivity should be matched with a given application. In the present study, fine discrimination among display configurations was of interest and two levels (i.e. High and Low) of task difficulty were manipulated by changing the size of the cell of the storage rack.

Therefore, the correlations of four different performance measures were evaluated further under the two task difficulty conditions. Tables 2 and 3 show correlation coefficients among performance measures under the high and low task difficulty conditions, respectively. While four different performance measures were highly correlated with each other under the high task difficulty condition, only the subjective rating was found to be highly correlated with the baseline performance measure of error rate under the low task difficulty condition (r=0.81).

Table 2. Correlation coefficients between performance measures under a high task difficulty condition.

	Error rate	Time to completion	Subjective ratings	Distance
Error rate	1			
Time to completion	0.91	1		
Subjective ratings	0.95	0.81	1	
Distance	0.95	0.93	0.92	1

Table 3. Correlation coefficients between performance measures under a low task difficulty condition.

	Error rate	Time to completion	Subjective ratings	Distance
Error rate	1			
Time to completion	- 0.29	1		
Subjective ratings	0.81	- 0.45	1	
Distance	0.28	- 0.05	0.37	1

Under the low level of task difficulty, variations in performance were not reflected by the time to completion and distance measures. It can be assumed that under the low level of task difficulty the operator has sufficient tolerance fit between the virtual object and the storage cell and therefore does not depend heavily on the depth information presented on the display screen in performing the task. Consequently, the operator can compensate for varying levels of display format and maintains adequate performance. The results imply that task difficulty can have a considerable effect on teleoperator performance and therefore the choice of performance measure techniques for teleoperational systems.

4. Discussions and Conclusion

In general, cares should be taken for the choice of appropriate performance measures in a particular application. One of the human factors issues in teleoperation studies is the comparison of control and display design options. Currently, a large number of techniques have been used as teleoperation performance measures. The present study

was performed to compare four different types of teleoperation performance measures by employing a simulated telerobotic task.

The Modified Cooper-Harper rating scale and distance measures were proposed as teleoperation performance measures based on subjective ratings and end-effector motions, respectively. These measures were compared to two commonly used performance measures, the error rate and time-to-completion measures. The distance measure showed close association with both the error rate and time-to-completion measures while the subjective measure showed close association with the error rate measure only. As task difficulty decreased, however, the close association was found only between the subjective measure and the error rate measure.

Based on the results of this study, the Modified Cooper-Harper rating scale, originally developed for aircraft handling tasks, can be applicable to telerobotic tasks. The subjective measure has a high correlation with the baseline measure of error rate and the close association between the two measures is not severely affected by the level of task difficulty. Subjective measures have practical advantages of being sensitive even in low task difficulty situations, requiring minimal instrumentation and operator training, and not degrading primary task performance. One of the potential disadvantages of using subjective assessments is the potential for confounding of mental and physiological load by operators (O'Donnell and Eggemeier, 1986). Since the task employed in the present study was manipulated by varying the display format and the control was fixed as an endpoint control using the SpaceBall, the possibility of confounding could be minimized. However, a further analysis seems to be required since even the same physical control may create different degree of load experienced by the operator observing different displays.

The distance measure is a simple measure of the end-effector motion controlled by the operator. The measure has a clear advantage of not requiring the definition of errors for a particular telerobotic task. The results of the present study indicate that the distance measure is highly correlated with the error rate and time-to-completion measures. However, the measure is not sensitive under the low level of task difficulty. One possible explanation for this result is that, in the present study, the distance measure was obtained by calculating the translational distance

moved by the end-effector, so that it did not account for the rotational movement of the end-effector. Further studies of a completed motion effort (including the rotational motion) of the end-effector will allow a better understanding of the robot's motion effort, as a teleoperation performance measure.

The results presented here indicate that the manipulation of task difficulty can have considerable effects on teleoperation performance and the effects vary with different performance measures. The choice of teleoperation performance measures should be made based on the task difficulty as well as the task itself in a particular application. The results revealed here also have implication for performance measures of various display and control systems including aircraft navigation displays and tactical displays.

References

Brouse, R. A. and McDonald, M. L. (1992). Using Tactile Information in Telerobotics. *IEEE Transactions on Systems, Man, and Cybernetics*, 22 (5), 1205-1210.

Casali, J. G. and Wierwille, W. W. (1982). A Sensitivity/Intrusion Comparison of Mental Workload Estimation Techniques Using a Flight Task Emphasizing Perceptual Piloting Activities. *Proceedings of the IEEE International Conference on Cybernetics and Society*, 598-602.

Casali, J. G. and Wierwille, W. W. (1983). Communications-Imposed Pilot Workload: A comparison of sixteen estimation techniques. *Proceedings of Second Ohio State University Symposium on Aviation Psychology*, 223-235.

Draper, J. V., Handel, S., and Hood, C. C. (1990). Fitts' s Task by Teleoperator: Movement time, velocity, and acceleration. *Proceeding of Human Factors Society 34th Annual Meeting, Human Factors Society*, Santa Monica, CA. 127-128.

Eggemeier, F. H. (1981). Current Issues in Subjective Assessment of Workload. *In Proceedings of the Human Factors Society 25th Annual Meeting.* Santa Monica, CA: Human Factors Society, 513-517.

Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology* (47), 381-391.

Hannaford, B., Wood, L., McAffee, D. A., & Zak, H. (1991). Performance Evaluation of a Six-Axis Generalized Force-Reflecting Teleoperator. *IEEE*

Transaction on System, Man, and Cybernetics, 21(3), 620-633.

Kim, W. S. & Stark, L. W. (1989). Cooperative Control of Visual Displays for Telemanipulation. *IEEE 1989 International Conference on Robotics and Automation*, vol. 3, 1327-1332.

Kim, W. S., Tendick, F., & Stark, L. (1987). Visual Enhancements in Pick-and-Place Tasks: Human operators controlling a simulated cylindrical manipulator, *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 5, 418-425, Oct.

Kugath, D. A. (1972). Experiments Evaluating Compliance and Force Feedback Effect on Manipulator Performance, General Electric Corp., NASA CR 128605, Philadelpia, PA.

MacKenzie, I. S. (1989). A Note on the Information-Theoretic Basis for Fitts' s Law. *Journal of Motor Behavior*, 21(3), 323-333.

Massimino, M. J. & Sheridan, T. B. (1994). Teleoperator Performance with Varying Force and Visual Feedback. *Human Factors*, 36(1), 145-157.

McGovern, D. E. (1974). Factors Affecting Control Allocation for Augmented Remote Manipulation. Doctoral dissertation, Design Division, Mechanical Engineering Department, Stanford University, Stanford, California.

McLean, G. F., Prescott, B., & Podhorodeski, R. (1994). Teleoperated System Performance Evaluation. *IEEE Transaction on System, Man, and Cybernetics*, 24(5), 796-803.

O' Donnell, R. D.and Eggemeier, F. T. (1986). Work Load Assessment Methodology. In K. Boff, L. Kaufman, and J. Thomas (eds.). *Handbook of perception and performance* (vol. II). New York: Wiley.

Park, S. H and Woldstad, J. C. (1998). Evaluation of Display Format and VisualEnhancement Cues for a Simulated Telerobotic Task. *1998 Researchers' Conference Proceedings*. Amarillo National Resource Center for Plutonium, Amarillo, Texas. 33-36, 1998.

Park, S. H and Woldstad, J. C. (2000). Multiple Two-Dimensional Displays as an Alternative to Three-Dimensional Displays in Tele-robotic Tasks. *Human Factors* (in press).

Rahimi, M. and Wierwille, W. W. (1982). Evaluation of the Sensitivity and Intrusion of Workload Estimation Techniques in Piloting Tasks Emphasizing Mediational Activity. *Proceedings of the IEEE International Con-*

ference on Cybernetics and Society, 593-597.

Sheridan, T. B. (1992). Telerobotics, Automation, and Human Supervisory Control. Cambridge, MA: MIT Press.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: HarperCollins Publisher.

Wickens, C. D. and Yeh, Y. Y.(1983). The Disassociation of Subjective Ratings and Performance: A multiple resources approach. *Proceedings of the Human Factors Society 27th Annual Meeting*. Santa Monica, CA: Human Factors Society, 224-248.

저자소개

◈ **박성하**

고려대학교 건축공학과 학사

The University of Texas, Arlington, Dept. of IE MSIE

Texas Tech University, Dept. of IE Ph. D.

현재: 한남대학교 산업공·기계공학부 조교수

관심분야: Human factors in teleoperation, aviation, and displays and controls. Industrial Ergonomics.

논문접수일 (Date Received): 2000/5/4

논문게재승인일(Date Accepted): 2000/8/10