Comparing Usability of User Interfaces for Robotic Telepresence

Federica Bazzano[1], Fabrizio Lamberti[1], Andrea Sanna[1], Gianluca Paravati[1] and Marco Gaspardone[2]
[1]*Politecnico di Torino, Dip. di Automatica e Informatica, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy*
[2]*TIM JOL Connected Robotics Applications LaB, Corso Montevecchio 71, 10129 Torino, Italy*

Keywords: Telepresence, Human-Robot Interaction (HRI), Augmented Reality, Robot Operating System (ROS).

Abstract: In the last years, robotic telepresence solutions have received a significant attention from both the commercial and academic worlds, due to their ability to allow people to feel physically present at a remote place and move in it. Operating a mobile robot with some autonomous capabilities from distance can enable a wide range of mass-market applications, encompassing teleconferencing, virtual tourism, etc. In these scenarios, the possibility to interact with the robot in a natural way becomes of crucial importance. The aim of this paper is to investigate, through a comparative analysis, the usability of two major approaches used today for controlling telepresence robots, i.e., keyboard and point-and-click video navigation. A control system featuring the above interfaces plus a combination of the two has been developed, and applied to the operation of a prototype telepresence robot in an office scenario. The system additionally includes functionalities found in many research and industry solutions, like map-based localization and "augmented" navigation. Then, a user study has been performed to assess the usability of the various control modalities for the execution of some navigation tasks in the considered context. The study provided precious indications to be possibly exploited for guiding next developments in the field.

1 INTRODUCTION

In the scientific world, the term "*telepresence*" was coined in 1980 by Marvin Minsky, an American cognitive scientist working in the area of artificial intelligence (AI) and co-founder of the AI laboratory of the Massachusetts Institute of Technology.

In an article published in 1980 (Minsky, 1980), he wrote: "[...] to convey the idea of remote control tools, scientists often use the words 'teleoperator' or 'telefactor'. I prefer to call this 'telepresence' [...]. It emphasizes the importance of high-quality sensory feedback and suggests future instruments that will feel and work so much like our own hands that we won't notice any significant difference".

By leveraging the above definition, it is possible to introduce the so-called "robotic telepresence", i.e., the area of robotics where the telepresence concept is applied to the control of a robot from a distance (Sheridan, 1995).

When dealing with the robotic telepresence area, two aspects are to be taken into particular account: the remote robot operation and the sensory feedback.

The former aspect refers to the fact that the robot and the operator do not share a common physical environment. The robot performs a given task in the environment based on instructions received from the remote human operator. The operator's mental strain resulting from performing this task in possibly complex operational conditions, coupled with his or her capability to respond to related demands, is generally termed mental or cognitive workload (Cain, 2007).

The latter aspect indicates that the human operator, while supervising the robot, receives information regarding its status and the surrounding environment, which has to be displayed in a way that lets he or she feel physically present at the remote site (Sheridan, 1992). The human's perception and understanding of this information is generally termed Situation Awareness (SA) (Endsley, 1988).

An effective design of teleoperation interfaces requires to identify the key elements to improve the operator's SA and lower his or her mental workload during the execution of the remote tasks, while keeping the interaction with the robot as simple as possible (De Barros and Linderman, 2009). The number and type of these elements directly translate into guidelines for designing effective techniques for human-robot interaction (HRI).

Examples of key elements regarding SA include

Bazzano F., Lamberti F., Sanna A., Paravati G. and Gaspardone M.
Comparing Usability of User Interfaces for Robotic Telepresence.
DOI: 10.5220/0006170300460054
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 46-54
ISBN: 978-989-758-229-5

video and range information, as well as robot's position and orientation (Nielsen and Goodrich, 2006). Each element has its own advantages and drawbacks. As a matter of example, video and range information is useful, e.g., to inform the operator about the presence of obstacles in the environment, but it is usually limited to the field of view of the sensors used to capture the video or to measure the distance from an object. Position and orientation are helpful to provide navigation reference to the human operator, but bandwidth limitations and communication delays could introduce critical mismatches in the frames of reference (Goodrich and Schultz, 2007).

Based on key elements mentioned above, telepresence interfaces are often broken down in two main categories: map-centric and video-centric (Keyes, 2007). In a map-centric interface, the map represents the most important input for the operator to supervise the navigation. In general, the map covers a wide region of the operator's display and all the relevant information is shown on it. In a video-centric interface, it is the video information that plays the dominant role.

Another key element that can significantly impact on the operator's mental workload is represented by the teleoperation paradigm used for remote robot navigation.

The main teleoperation paradigms have been historically classified into four categories, namely, direct, multimodal/multisensor, supervisory, and "novel" (Fong and Thorpe, 2001). Direct interfaces include all the "traditional" hand-controlled devices, like mouse, keyboard, joystick, etc. Multimodal/multisensor interfaces offer multiple control modes combined according to temporal and contextual constraints in order to allow their interpretation. Examples are voice and gesture commands combined with traditional inputs (Stiefelhagen et al., 2007). Supervisory control interfaces are designed for robots with some level of autonomy. These interfaces provide methods for reviewing results, so that the operator can monitor and identify execution anomalies. Novel interfaces use unconventional input methods, e.g., based on brainwave and muscle movement monitoring (hence the name, basically meaning not included in the previous categories).

In the last years, researchers from both academy and industry proposed a number of interfaces for robots' teleoperation, often obtained by merging more than one of the approaches reported above. Thus, none of these solutions has emerged yet as the ultimate approach to robotic telepresence.

By leveraging the above observations, this paper reports on the activities that have been carried out at Politecnico di Torino and at TIM JOL Connected Robotics Applications LaB (CRAB) to investigate, through a comparative analysis, the usability of two major approaches that are used today in the industry. The analysis has been performed by working with a telepresence robot that has been exploited so far in cultural heritage scenarios (Giuliano et al., 2015).

A user study was carried out with several volunteers, who were asked to carry out some navigation tasks in a office environment by working with a teleoperation solution integrating both map and live video "augmented" with navigation cues. Volunteers were invited to control the robot by using a direct and a supervisory approach (as well as a combination of the two) and to judge the various control modalities from the point of view of usability. Data about time and number of interactions required to complete the tasks were additionally recorded in order to measure the effectiveness of the various modalities.

The rest of the paper is organized as follows. In Section 2, relevant works in the area of telepresence robots are reviewed. In Section 3, the robot considered in the study is described. Section 4 provides an overview of the overall control system, and reports the details of the navigation modalities that have studied. Section 5 introduces the methodology that has been adopted to perform the experimental tests and discusses results obtained. Lastly, Section 6 concludes the paper by providing possible directions for future research activities in this field.

2 RELATED WORK

Many works in the HRI domain investigated the design of interfaces for remote robotics applications. Examples are reported in the academic literature and are provided by commercial solutions as well.

For instance, in (Lewis et al., 2014) the map-centric interface designed by the iRobot Corporation for the Ava 500 is presented. The interface allows the operator to define the target location for the robot by clicking on a 2D map. Similarly, in (Drury et al., 2007), the map-centric system developed by the MITRE Corporation is illustrated. In this system, multiple robots are used to build up a map of the environment. The devised interface allows the operator to switch the navigation commands among the various robots, by defining their target location as 2D coordinates. Small windows containing the video stream received from the robots appear under the map. In (Nielsen et al., 2004), the activities of the Idaho National Laboratory aimed to the creation of a semantic map-centric interface that combines information from the environment with a 3D virtual map are presented. The map is augmented by icons or symbols defined by the operator during a preliminary exploration of the environment, which are meant to provide a meaning for given places and objects.

A limitation of this kind of interfaces could be represented by the map itself. In fact, faulty sensors or dynamic objects in the environment could lead to the creation of incorrect maps, which would reduce the operator's SA.

While the above solutions were experimented, other works focused on the design of video-centric interfaces. As a matter of example, in (Zalud, 2006), a video-centric interface with 3D-display capabilities is illustrated. The remote operator wears virtual reality googles that show information about the robot and the environment overlapped to the video stream. The operator can change the point of view of the robot's camera based on where he or she is looking at, and can guide the robot throughout the environment by means of a two-hand joystick. Two severe drawbacks of this interface are the immersion sickness due to the use of a wearable display and the requirement about the orientation of the operator's head (that needs to be enforced in order to maintain the robot's camera properly centred). A different video-centric interface has been developed by the Swarthmore College (Maxwell et al., 2003). In this case, researchers used augmented reality information overlapped to the video stream in order to show the distance of the robot from an obstacle and the pan-tilt-zoom indicators. Others information, such as the infrared and sonar distance data, location on the map and camera orientation are positioned around the screen area. Navigation inputs are passed to the robot by using the keyboard or by clicking on control buttons in the interface. The main limitation of this interface is the way sensor data are displayed. Visualization could be improved by merging them with the map information, in order to allow the operator to better understand where the robot is actually located (Keyes, 2007).

Works reviewed above have been selected as representative examples of the developments made by the academy. From a commercial perspective, it can be observed that telepresence robots often rely on video-centric interfaces. They mostly include also a map, and generally exploit direct or supervisory teleoperation. Some examples are the Beam Smart Presence Robot (Neustaedter et al., 2016), the VGo Robot (Tsui et al., 2011), the Padbot Robot (Lee et al., 2015) and the Ra.Ro Robot[1]. Their interfaces use traditional technologies to let the operator define the coordinates of the target destination on a map, while supervising the robot through a live video stream. The latter robot recently integrated a further navigation method called "smart drive", which allows the operator to guide the robot by pointing and clicking a specific location directly on the video stream, thus obtaining the 3D coordinates to be passed on to the robot.

Based on the short review above, it can be observed that the panorama of interfaces available for remotely operating telepresence robots is quite heterogeneous, and data about suitability of a particular implementation for a specific scenario are actually required in order to properly support next advancements in the field. Some activities in this direction have been already carried out. For instance, in (Nielsen and Goodrich, 2006), map-centric and video-centric interfaces are compared, by studying three configurations, i.e., map-only, video-only, and map plus video. The authors concluded their study by stating that, when video and map information are integrated, they tend to complement each other and to improve overall performances.

The goal of the present paper is to build on results reported in (Nielsen and Goodrich, 2006) to study a richer scenario in which a mixed map & video-based control system is considered, and two major robot's navigation interfaces used in recent telepresence solutions available on the market as well as their combination are compared from a usability perspective.

3 ROBOTIC PLATFORM

This section briefly describes the robotic platform that has considered in this study, named Virgil, by illustrating its hardware and software features.

Virgil is a telepresence robot developed in the frame of the collaboration between TIM JOL Connected Robotics Applications LaB and Politecnico di Torino. It was originally designed to be used in cultural heritage scenarios and allow museum visitors to explore areas generally closed to the public (Fig.1). The tour guide could take control of the robot remotely and teleoperate it in restricted areas by displaying a live video stream to the visitors at the same time.

The robot is based on a wheeled mobile platform equipped with a pan-tilt camera and a laser sensor. The former device allows the tour guide to remotely move the focus on the area of interest without moving the robot, whereas the latter device allows the robot to avoid possible obstacles present in the environment. It weights about 14 Kg and is 120 cm tall. It is equipped with a Li-Fe 12V battery, which provides it with an autonomy of approximately 4 hours

[1] http://www.nuzoo.it/en/

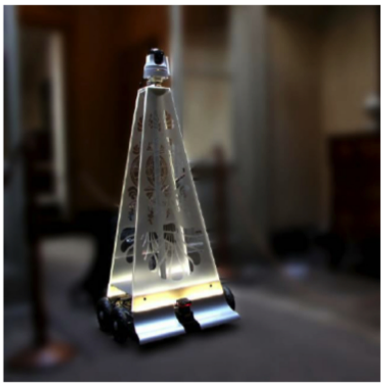


Figure 1: Virgil.

and a maximum velocity of 1 m/s. Virgil is capable to navigate throughout a given environment in an autonomous way by using its local and global path planning functionalities, which rely on a map of the environment created in a preliminary exploration phase. Algorithms are executed on the Robot Operating System (ROS)-based platform for cloud robotics created by TIM.

With respect to the original setup, in this study the robot has been modified by adding a tablet device, whose camera replaces the original sensor. The screen is used to display at the remote site the face of the operator, thus enhancing the sense of presence.

4 NAVIGATION MODALITIES

In this work, two navigation modalities have been studied, which differ in the way the operator can control the robot and in the type of feedback returned.

In both the modalities, the interface displays a large video window, in which the live stream from the robot's camera is shown (a smaller window shows the video captured by a local webcam, which is displayed on the remote tablet mounted on the robot).

In the first modality, later referred to as *keyboard teleoperation*, the operator manually guides the robot throughout the environment by means of directions keys. In the second interface, named *point-and-click video navigation*, the operator either issues step-by-step direction commands or specifies a target destination by clicking it on the video stream received by the robot's camera; depending of the command issued, the robot executes the particular commands or manages to reach the destination in an autonomous way.

In the second modality, the path that is being followed by the robot to reach the clicked target is additionally shown, overlapped to the video stream. Right below the video window there is a colored bar, which is used to display the distance from obstacles. The bar is split in three regions, that depict obstacles in front, to the left or the right of the robot. Bar color changes from green to red based on the actual measurements of the laser sensor.

The interface also includes a map, where the position and orientation of the robot is shown in real time. In particular, a yellow triangle is used to indicate the robot's location and orientation. In the second modality, a green marker is used to represent the clicked destination and the orientation the robot will assume once it will reach it. It is worth noting that navigation algorithms actually work on a different map, which was created by using the robot's laser sensor and applying a Simultaneous Localization and Mapping (SLAM) strategy.

In the present study, the focus has been kept on robot's navigation. Hence, the possibility to pan-tilt the camera has not been considered.

The two modalities have been implemented in a Web application, which communicates with the robot and the cloud robotics platform that hosts the navigation algorithms via `roslibjs`, a JavaScript-based library for using ROS on the Web (Toris et al., 2015). Details of two modalities and screenshot of the corresponding interfaces are reported in the next subsections.

4.1 Keyboard Teleoperation

As said, this configuration allows the user to move the robot into the remote environment with the keyboard keys. The up and down arrow keys are associated with a ROS command that changes robot's linear velocity, making the robot move forward or backward in the environment. The left and right keys change angular velocity, twisting the robot accordingly. When pressed together, the above keys can be used to make the robot move in the given direction while turning left or right. Command issued and current robot's direction are displayed on the video stream using augmented reality arrows.

The same commands can be issued by clicking the corresponding icons. In this case, two icons are used to control the simultaneous use of direction and orientation commands. The operator could also change the robot's speed by using the sliders in the bottom part of the interface shown in Fig. 2 (not used in the present study to enable comparability of results, since arbitrary speed values are not supported by the path planning algorithms adopted).

In this modality the feedback provided to the operator is represented by the live video stream and by the map. The robot exploits a local path planning algorithm to navigate the environment. Hence, informa-

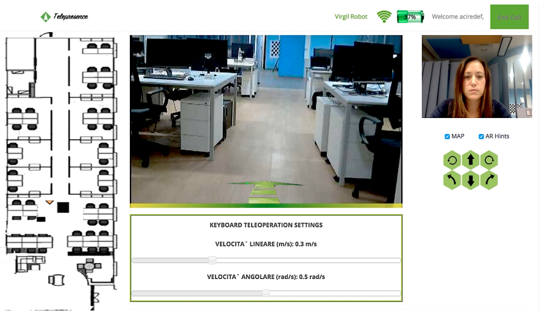


Figure 2: Keyboard teleoperation interface.

Figure 3: Point-and-click video navigation.

tion regarding the distance from obstacles is shown in the colored bar below the video stream.

4.2 Point-and-click Video Navigation

This configuration assumes that the operator defines a target destination for the robot by simply clicking it in the live video stream window, as shown in Fig. 3.

Since the intrinsic parameters of the camera and (fixed) pan-tilt configuration are known, the coordinates of the pixel clicked by the operator can be converted to a point on the map by using ray-tracing. This latter point is passed as a goal to the global path planning algorithm, which is designed to move the robot towards the target destination at a constant speed by avoiding both fixed and moving obstacles.

With respect to previous modality, at any point in time the robot knows the path elaborated by the planning algorithm to reach the goal. This path is overlapped to the video stream in augmented reality, as shown in Fig. 4. This way, operator's SA is further increased with respect to previous modality.

It is worth observing that, when the robot is close to an obstacle that occupies the whole field of view of the onboard camera, the operator is not able to find any point to click. Hence, in order to make this modality comparable to the keyboard-based one, a direct control method still based on clicking on the video stream has also been implemented. With this method, the operator can click on several active areas on the edges and corners of the video window (like in

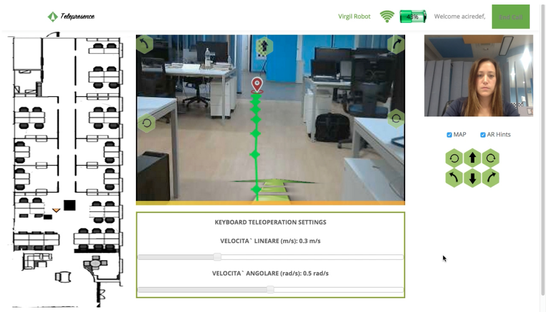


Figure 4: Point-and-click video navigation interface.

Google Street View[2]) to directly control robot's linear or angular velocity. When such commands are issued, goals that might have been set for autonomous navigation are cleared. A new goal could then be specified once the target destination or an intermediate location are visible again in the camera's field of view.

5 EXPERIMENTAL RESULTS

As anticipated, the goal of this paper is to perform an evaluation of the *keyboard teleoperation* (later abbreviated, *keyboard*, or *K*) and *point-and-click video navigation* (abbreviated *point-and-click*, or *P*) interfaces and the combination of the two (*combined*, or *C*) from the perspectives of usability and effectiveness.

Evaluation was carried out based on objective and subjective observations collected through a user study. The study involved 12 participants (9 males and 3 females, aged between 24 to 27), selected from the students of Politecnico di Torino. According to declarations collected, 80% of them had already used keyboard-based interfaces to issue direction commands (e.g., in video-games), and 50% had previous experience with interfaces based on point-and-click.

Each participants was invited to control the robot by using all the above interfaces in order to carry out three navigation tasks, referred to as as *T1 - Reach the column*, *T2 - Reach the room*, and *T3 - Enter/exit the room*. Such tasks have been specifically designed to test the suitability of the various interfaces in the possible scenarios the robot could be involved into when used in an office scenario.

In particular, *T1* was meant to test the interfaces when guiding the robot to a destination that is outside the camera's field of view. *T2* was meant to test the interfaces when obstacles are to be avoided. Lastly, *T3* was designed to test the interfaces when guiding the robot in constrained spaces.

[2]https://www.google.com/streetview/

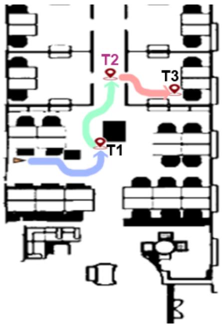


Figure 5: Map of the environment considered in the experiments, initial location of the robot, destinations to be reached in the tasks and possible paths.

Experiments were performed in the Graphics and Intelligent Systems (GRAINS) Research Laboratory in Turin on a laptop computer equipped with Ubuntu 14.04 LTS (Trusty Tahr) operating system, a monitor with resolution 1920×1080 and an optical mouse as input device.

After a brief training, participants were invited to perform the three tasks in a sequence by using all the interfaces. The interface to be used was chosen in a random order so that to limit the effect of learning.

The robot is initially standing in the open space of JOL laboratories in Turin. Its position is indicated by the yellow triangle marker in the map shown in Fig. 5. This choice has been made to ensure that the camera cannot frame the destination to be reached in the first task. The participant firstly has to guide the robot close to a column identified by label *T1* in the map. The path to be possibly followed is shown by the blue arrow. Afterwards, the robot needs to be moved to a second checkpoint (identified by label *T2*) located in front of a particular meeting room by avoiding the column. Possible path is shown in green. Finally, the participant has to guide the robot into the room, bring it close to a desk (indicated by label *T3*), turn around and exit the room. Path is marked in red in the map.

During each experiment, quantitative data about time required to complete the tasks and number of interactions (key presses and/or mouse clicks, depending on the interface considered) were recorded. At the end, each participant was asked to fill in a usability questionnaire[3] split in three parts.

The first part was created by considering the Nielsen Attributes of Usability (NAU) (Nielsen, 1994). NAU requested participants to evaluate five statements referring to as many usability factors, namely, Learnability, Efficiency, Memorability, Errors and Satisfaction by expressing their agreement on a 4-point Likert scale.

[3]http://goo.gl/NzANxb

The second part was created by considering the Subjective Assessment of Speech System Interfaces (SASSI) methodology (Hone and Graham, 2000) and adapting it to let participants judge the user experience with the given interaction means. The (adapted) SASSI questionnaire requested participants to evaluate 6 statements referring to as many usability factors, i.e., System Response Accuracy (SRA), Likeability (LIKE), Cognitive Demand (CD), Annoyance (AN), Habitability (HAB) and Speed (SPE) by expressing their agreement on a 4-point Likert scale. It is worth observing that other tools specifically tailored to the evaluation of mental workload could be exploited as well (Rubio et al., 2004), (Kiselev and Loutfi, 2012).

The third part asked participants to rank, for each task, the experience made with the three interfaces by providing their judgment both for the three individual tasks as well as for the whole experiment.

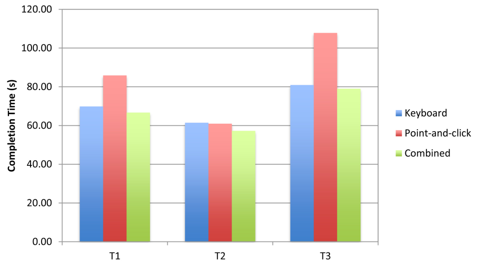
Results obtained in terms of completion time as well as number of interactions required to complete the tasks are reported in Fig. 6. At first sight, it appears that, in *T1* and *T3*, completion times obtained with the *keyboard* and the *combined* interfaces were largely lower than those experimented with the *point-and-click* one (Fig. 6(a)). Statistical significance was confirmed by running paired samples t-test (P largely lower than 0.05). Results for *T2* were not statistically significant. Average values for the number of interactions (Fig. 6(b)) indicate that, for all the tasks, the *point-and-click* and the *combined* interfaces required a lower number of interactions compared to the *keyboard* one (this number was much lower in the case of *T2*).

Lower completion time and reduced number of interactions for the *keyboard* and the *combined* interfaces are observed also when summing up results obtained for the three tasks, i.e., considering them altogether as a single experiment.

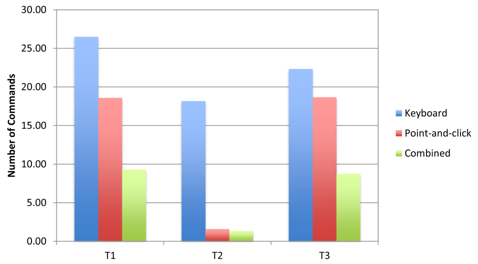
Results obtained with the subjective evaluation based on the NAU methodology (Fig. 8) appear to describe an almost comparable situation overall, since participants judged the *keyboard* and *combined* interfaces more usable than the *point-and-click* one for what it concerns the five factors considered.

Similar considerations can be made for five out of the six usability factors of the (adapted) SASSI methodology. In fact, as shown in Fig. 9, the *point-and-click* interface performed better than the other ones only in terms of Annoyance, i.e., in terms of how much the interface was judged repetitive and boring.

Results regarding users' preferences in using the three interfaces to carry out individual tasks as well as the whole experiment are shown in Fig. 7. Considering overall rankings, it appears that the favorite



(a)



(b)

Figure 6: Results in terms of (a) completion time and (b) number of interactions required to complete the tasks with the considered interfaces. Bar heights report average values (lower is better).

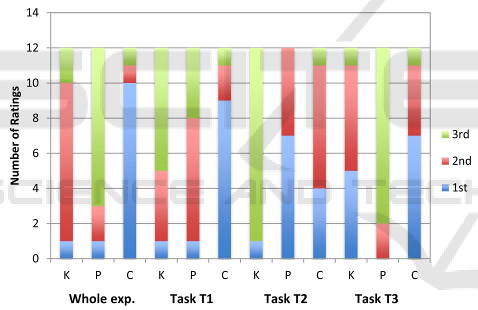


Figure 7: Number of times the *keyboard* teleoperation (K), *point-and-click video navigation* (P) and *combined* (C) interfaces have been ranked 1st, 2nd and 3rd for the execution of the whole experiment and individual tasks.

interface is the *combined* one. When individual tasks are considered, it could be observed that the *combined* interface is the one that was preferred for performing *T1* and *T3* (see in particular blue and red columns in Fig. 7, where the number of times a given interface has been ranked 1st or 2nd is showed).

Concerning the users' gender classes, statistically significant differences in terms of completion time were found using Anova tests (0.05 significance level). Females were faster than males in executing T3 with the *point-and-click* interface and T1 with the *combined* one. As for users' previous experience with keyboard-based interfaces, the task completion time of those who indicated an everyday usage frequency was lower than those who stated to be used to work with such kind of interfaces once a week. A similar consideration holds also for the number of interactions required to complete the tasks. Results for prior knowledge about point-and-click interfaces and subjective observations were not statistically significant.

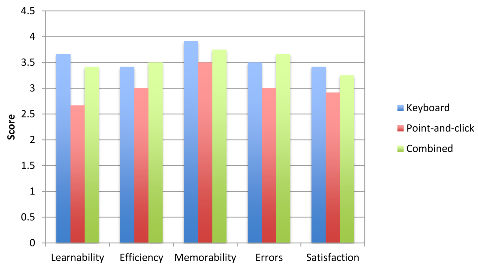


Figure 8: Results concerning the usability of the three interfaces for the whole experiment based on NAU factors. Bar heights report average values (higher is better).

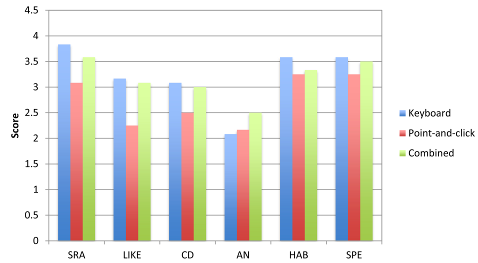


Figure 9: Results concerning the usability of the three interfaces for the whole experiment based on (adapted) SASSI methodology. Bar heights report average values (higher is better).

Based on the feedback gathered in the tests, preference seems to be mainly motivated by to fact that, as it could be largely expected, participants were allowed to switch between the two interfaces when needed, thus benefiting from the advantages of both of them. This behavior can be appreciated also considering results discussed above. However, by combining these results with those regarding users' interaction, it is evident that, in the execution of the above tasks, participants mainly used the *keyboard* interface. The *point-and-click* interface was largely preferred in the execution of *T2* and slightly preferred in the execution of *T1*. This is reasonably due to the fact that, in these tasks, robot was requested to move over long distances. In these situations, autonomous navigation capabilities could effectively limit the participants' workload, who simply need to click on the destination to reach (by possibly specifying intermediate waypoints or adjusting the path by issuing directional commands using the active regions on the borders and corners of the video window). It is also worth adding that most of the concerns regarding the use of the *point-and-click* interface were related to specific field of view of the camera, which often limited the possibility for the operator to immediately spot the destination to click. This limitation could be addressed by adding the possibility for the various interfaces to control also the pan-tilt of the camera.

6 CONCLUSIONS

In this work, a comparative analysis of user interfaces for robotic telepresence applications was presented. The evaluation focused on two main control modalities used today in most research and commercial solutions, i.e., keyboard and point-and-click video navigation. The combination of the two was also considered.

Experimental results obtained through a user study provided precious indications about user experience with the three interfaces, both in objective and subjective terms. In particular, results obtained with the objective evaluation showed that, through the combined interface, time required to perform a complex task and commands to be issued can be significantly reduced. Similarly, results obtained with the subjective evaluation suggested that the favorite interface was the combined one. However, based on the feedback gathered in the tests, it could be observed that users' preference for the combined interface was due to fact that they were allowed to switch between the two interfaces when needed, thus benefiting from the advantages of both of them. By digging more in details in performances obtained and preferences expressed in the execution of specific navigation operations, it was found that the keyboard-based interface actually provided significant advantages when accurate control was needed, whereas point-and-click video navigation was more effective when robots autonomous navigation capabilities could be exploited.

Future works will be aimed to address concerns regarding the possibility to control the pan-tilt of the robot's camera in the point-and-click video navigation interface, which should translate into even better performances for the combination of the two interfaces and for next-generation robotic teleoperation solutions.

REFERENCES

Cain, B. (2007). A review of the mental workload literature. Technical report, DTIC Document.

De Barros, P. G. and Linderman, R. W. (2009). A survey of user interfaces for robot teleoperation. *Tech. Rep. Series, WPI-CS-TR-09-12*.

Drury, J. L., Keyes, B., and Yanco, H. A. (2007). Lassoing hri: analyzing situation awareness in map-centric and video-centric interfaces. In *Human-Robot Interaction (HRI), 2nd ACM/IEEE Int. Conf. on*, pages 279–286. IEEE.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, volume 32, pages 97–101. SAGE Publications.

Fong, T. and Thorpe, C. (2001). Vehicle teleoperation interfaces. *Autonomous robots*, 11(1):9–18.

Giuliano, L., Ng, M. E. K., Lupetti, M. L., and Germak, C. (2015). Virgil, robot for museum experience: study on the opportunity given by robot capability to integrate the actual museum visit. In *Intelligent Technologies for Interactive Entertainment (INTETAIN), 7th Int. Conf. on*, pages 222–223. IEEE.

Goodrich, M. A. and Schultz, A. C. (2007). Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3):203–275.

Hone, K. S. and Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Eng.*, 6(3&4):287–303.

Keyes, B. (2007). Evolution of a telepresence robot interface. *Unpublished master's thesis. University of Massachusetts, Lowell*, 7.

Kiselev, A. and Loutfi, A. (2012). Using a mental workload index as a measure of usability of a user interface for social robotic telepresence. In *2nd Workshop of Social Robotic Telepresence in Conjunction with IEEE International Symposium on Robot and Human Interactive Communication 2012*.

Lee, H., Choi, J. J., and Kwak, S. S. (2015). A social agent, or a medium?: The impact of anthropomorphism of telepresence robot's sound interface on perceived copresence, telepresence and social presence. In *Proc. of the 7th Int. workshops on the convergent Research Society among Humanities, Sociology, Science, and Technology*, pages 19–22.

Lewis, T., Drury, J., and Beltz, B. (2014). Evaluating mobile remote presence (mrp) robots. In *Proc. 18th Int. Conf. on Supporting Group Work*, pages 302–305. ACM.

Maxwell, B. A., Ward, N., and Heckel, F. (2003). A human-robot interface for urban search and rescue. *AAAI Mobile Robot Competition*, 3(1).

Minsky, M. (1980). Telepresence. *Omni*, pages 45–51.

Neustaedter, C., Venolia, G., Procyk, J., and Hawkins, D. (2016). To beam or not to beam: A study of remote telepresence attendance at an academic conference. In *Proc. of the 19th ACM Conf. on Computer-Supported Cooperative Work & Social Computing*, pages 418–431. ACM.

Nielsen, C. W. and Goodrich, M. A. (2006). Comparing the usefulness of video and map information in navigation tasks. In *Proc. of the 1st ACM SIGCHI/SIGART conf. on Human-robot interaction*, pages 95–101. ACM.

Nielsen, C. W., Ricks, B., Goodrich, M. A., Bruemmer, D., Few, D., and Few, M. (2004). Snapshots for semantic maps. In *Systems, Man and Cybernetics, IEEE Int. Conf. on*, volume 3, pages 2853–2858. IEEE.

Nielsen, J. (1994). *Usability engineering*. Elsevier.

Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Applied Psychology*, 53(1):61–86.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press.

Sheridan, T. B. (1995). Teleoperation, telerobotics and telepresence: A progress report. *Control Engineering Practice*, 3(2):205–214.

Stiefelhagen, R., Ekenel, H. K., Fugen, C., Gieselmann, P., Holzapfel, H., Kraft, F., Nickel, K., Voit, M., and Waibel, A. (2007). Enabling multimodal human–robot interaction for the karlsruhe humanoid robot. *IEEE Transactions on Robotics*, 23(5):840–851.

Toris, R., Kammerl, J., Lu, D. V., Lee, J., Jenkins, O. C., Osentoski, S., Wills, M., and Chernova, S. (2015). Robot web tools: Efficient messaging for cloud robotics. In *Intelligent Robots and Systems (IROS), IEEE/RSJ Int. Conf. on*, pages 4530–4537. IEEE.

Tsui, K. M., Desai, M., Yanco, H. A., and Uhlik, C. (2011). Exploring use cases for telepresence robots. In *6th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, pages 11–18. IEEE.

Zalud, L. (2006). Argos-system for heterogeneous mobile robot teleoperation. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 211–216. IEEE.