User-Centered Evaluation of the Virtual Binocular Interface

Donald Glowinski[1], Maurizio Mancini[1], Paolo Coletta[1], Simone Ghisio[1], Carlo Chiorri[2], Antonio Camurri[1], and Gualtiero Volpe[1]

[1] InfoMus Lab, DIST, University of Genova, Italy
{donald,maurizio,ghisio}@infomus.org,
{paolo.coletta,antonio.camurri,gualtiero.volpe}@unige.it
[2] DiSA - Department of Anthropological Sciences
Psychology Unit, University of Genova, Italy
carlo.chiorri@unige.it

Abstract. This paper describes a full-body pointing interface based on the mimicking of the use of binoculars, the Virtual Binocular Interface. This interface is a component of the interactive installation “Viaggiatori di Sguardo”, located at Palazzo Ducale, Genova, Italy, and visited by more than 5,000 visitors so far. This paper focuses on the evaluation of such an interface.

Keywords: Virtual Reality, Interactive Museum Applications and Guides, Novel Interaction Technologies.

1 Introduction

In this paper we propose the evaluation of a full-body pointing interface: the *Virtual Binocular Interface*. The permanent interactive museum installation “Viaggiatori di Sguardo”, opened on December 2009 at Palazzo Ducale, Genova, Italy, allows visitors to explore and interact with an audiovisual content by mimicking the use of binoculars. Via this interface, users can start a virtual journey to discover the monumental buildings “Palazzi dei Rolli” (UNESCO Treasure) of the Italian city of Genova.

In such context, the role of the user can resemble that of an *explorer* or a *traveller*, *viaggiatore* in Italian. The metaphors of *journey*, *exploration* and *travelling* led us to conceive, design, and implement the Virtual Binocular Interface to explore, in an ecological way, audiovisual content.

2 “Viaggiatori di Sguardo”: Overview and Background

The installation “Viaggiatori di sguardo”, designed and developed by Casa Paganini - InfoMus Research Centre, University of Genova, in collaboration with Palazzo Ducale, Fondazione per la Cultura of Genova, is a sensitive environment located in a room available to the public of tourists and visitors of Palazzo

A. Camurri, C. Costa, and G. Volpe (Eds.): INTETAIN 2011, LNICST 78, pp. 63–72, 2012.

Ducale in Genova, Italy. The goal of the installation is to present to visitors the UNESCO treasure of "Palazzi dei Rolli", a number of magnificent monumental buildings located in the city.

Fig. 1. The "Viaggiatori di Sguardo" permanent installation at Palazzo Ducale

Figure 1 shows the installation paradigm: a large screen (7 meters) that can be explored by several users by via the *Virtual Binocular Interface*. Below the large screen, smaller LCD screens are installed to provide brief textual information on the artwork that can be explored. The user, being a traveler on such cultural heritage, has at her disposal the Virtual Binocular: when she mimics with both hands the gesture of raising and pointing a binocular, she is enabled to zoom in the available cultural heritage audiovisual content, as shown on the left side of Figure 2.

Such gesture is functional and ecological in the context of the exploration of unknown places, since the user feels at her ease in such interaction. In several other cases of full body interfaces designed for interaction with content, it has been reported by users a varying extent of shame or embarrassment, mostly due to the need to perform unnatural movements in presence of others [9]. The proposed interface may represent a solution to this and other problems.

On the right side of Figure 2 one of the mechanisms available to navigate the monumental buildings is shown: in some spots of the image explored by the user, the contour of the binocular shape projected on the screen changes colour: this means that if the user continues pointing to that spot for a few seconds then she will enter deeper in the selected location (e.g., in a room or in a painting inside the monumental building).

3 Evaluation

This paper focuses on the evaluation of the Virtual Binocular Interface. A first pilot evaluation was conducted by adapting the cyclical multi-direction pointing

Fig. 2. Illustration of the Virtual Binocular Interface interaction paradigm. On the left: the user mimics the gesture of pointing with a binocular; the shape of a binocular view is projected on the screen allowing the user to zoom on the displayed picture. On the right: the binocular shape changes colour, that is, if the user continues pointing to that spot for a few seconds then she will enter deeper in the selected location.

task paradigm of ISO 9241-9 [12,11]. By mimicking the use of a binocular, the user had to reach, one at a time, 8 visual points displayed in a circle. The binocular posture and movement were captured by infrared video cameras and the EyesWeb XMI open software platform [2] was used to develop the binocular interface. Preliminary results showed a relatively high difference between the movement times to reach targets in the vertical direction with respect to the horizontal one [1]. We hypothesized that the difference may result from the application of different algorithms to track user movements on the vertical and horizontal planes: the former one is based on optical flow, whereas the latter relies on geometric features.

We decided to investigate such difference in detail by adapting the serial Fitts' paradigm for pointing task in a new experiment [5]. The serial paradigm was preferred to assess interface usability by focusing on performance efficiency and comfort in repetitive tasks. Specifically, we aimed at assessing the effects of target orientation, type of algorithm and difficulty of the task on movement time, controlling for task error rate. The evaluation of participants' performance was then refined by considering two original features (*geometry entropy* and *directness indexes*) to obtain a more qualitative description of the task.

3.1 Set-Up

The Virtual Binocular Interface is implemented by analyzing in real-time the movements of the participants in a room through infrared cameras. The EyesWeb platform [2] is used to analyze in real-time the video signal and identify the relevant gestures of the participant.

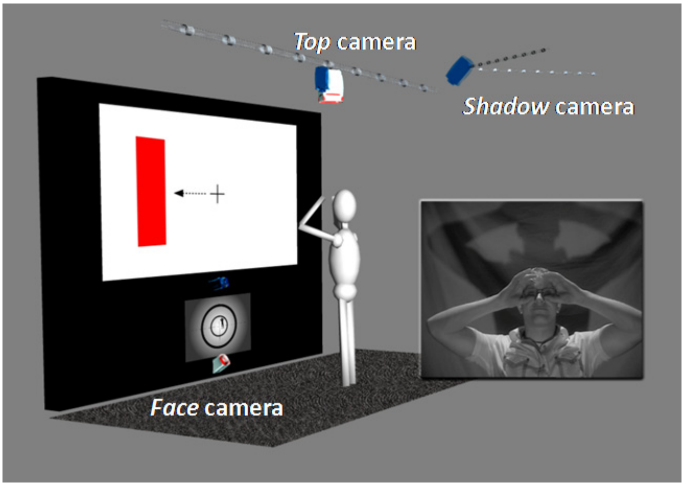

Fig. 3. Setup of the experiment

The participant is monitored by three infrared video cameras, the *top*, the *shadow*, and the *face* cameras (see Figure 3):

1. the top camera is installed above the user, to map from the vertical perspective the position on the floor of the head of the user.
2. the shadow camera is placed above the projection screen, looking toward the wall behind and above the user, where the shadow of the visitor upper-body part is projected by an IR light placed on the floor in front of the user. The shadow is analyzed to identify the binocular posture.
3. the face camera is installed below the projection screen, near the IR light, toward the user face, with an elevation of about 45 degrees. It is used to analyze the vertical movement of the user's face and control the up/down movement of the binocular.

The analysis of up-down movements is based on optical flow techniques applied to the face. Starting from the face camera, a region of interest corresponding to the user face is segmented and tracked. Then two optical flow algorithms have been selected and evaluated: the HornSchunck and the LucasKanade [6,7]. Left-right movement are estimated using visual information from the face camera or the top camera. In the first case, blob tracking techniques are used to detect the user rotation around the vertical axis. In the second case, the Lucas-Kanade optical flow is again used for tracking displacement along the horizontal axis. This leads to four different solutions to implement the Virtual Binocular Interface: Horn-Schunck (Algo1) and Lucas Kanade (Algo2) for up-down movements, blob tracking (Algo3) and Lucas Kanade optical flow (Algo4) for left-right movements. We tested all four solutions using a standard experimental procedure in order to select the best one, which has been chosen for the public permanent installation "Viaggiatori di Sguardo".

3.2 Procedure

Participants. Twenty subjects (10 male, 10 female, age 23-50) were recruited and participated on a voluntary basis. All subjects were healthy and had normal or corrected-to-normal vision.

Task and Design. The serial tapping task developed by Fitts [5,11] was adapted to perform our test. Participants were instructed to alternately reach as quickly and accurately as possible between two targets shown on a screen in front of them of width W at a distance D both in the horizontal and in the vertical axis. The target was presented as a red rectangle on a screen of 240 x 180 cm with a resolution of 720 x 576 pixel (see Figure 3). Standing position in front of the screen was fixed by the area delimited by the top camera view (around 1.5 m from the screen). Following classic Fitts' law studies [5], movement difficulty was manipulated by varying target width (W) and distance (D). W ranged from 25 to 40 pixels. D ranged from 216 to 576 pixels. Four W/D combinations were generated to cover four indexes of difficulty (ID) that represent typical tasks performed by users in the Viaggiatori di Sguardo installation (e.g., performing short distance between large targets or performing long distance between small targets). For each ID, the participant had to go back and forth between the two targets for 10 trials. To control for order and sequence effects, the order of ID differed for each subject according to a balanced Latin Square. The number of trials per participant for testing the Virtual Binocular was: Orientation (2) x Algorithm (2) x ID (4) x Trials (10) = 160 trials.

Procedure and Instructions. The experiment started when the participant began mimicking the handling and use of the Virtual Binocular (i.e. raise both hands at the level of her/his eyes). The participants could then move the crosshair displayed in the center of the screen through their upper-body part movements and reached the targets one at a time, dwelling over each of them for at least 0.75s. A maximum time of 3 seconds was allowed to reach each target. If a target was missed, participants were instructed not to try to correct the error, but to continue to the next target. Before starting the experiment, a sequence of warm-up trials, which consisted in reaching 4 targets twice without error, was performed. During the experimental session, resting duration was freely decided by participants according to the level of their muscular/mental workload that resulted from the task demands. After the experiment, information about the subject background and verbal accounts of pointing strategies were collected through questionnaires developed by [3]. The overall duration of the experiment was about 20 minutes. This relatively brief duration with respect to usual experiment in this field [11] was motivated by effort required to stand in binocular posture.

3.3 Results

The experiment yielded 3200 data points (20 subjects x 2 orientations x 2 algorithms x 4 ID x 10 trials). In each ID condition, outliers of more than 3 SD from the mean were not included in the movement time (MT) analysis. These removals left 3196 data points.

Mean Movement Time and Error Rate. A mixed three-way ANCOVA was performed on the Movement Time (MT) of participants with algorithm (Algo), Index of Difficulty (ID) and Orientation as factors and error rate as covariate. Main significant effects were found for ID ($F3, 2527 = 51.605$, $p < 0.001$), Orientation ($F2, 2527 = 3.876$, $p < 0.05$) and ID x Algo interaction ($F6, 2527 = 4.061$, $p < 0.001$). Bonferroni-corrected post-hoc comparisons revealed that Movement Time (MT) values for the targets located on vertical axis are significant lower than the targets located on horizontal axis confirming preliminary results. The MT values corresponding to higher Index of Difficulties were also significantly higher. In addition, a polynomial contrast of order 1 (linear) identified a positive trend of MT means for the four IDs. Post-Hoc comparisons of ID x Algo interaction effects showed that when the Lucas-Kanade algorithm is used for monitoring the participants along the horizontal axis (Algo 4), levels of difficulty (IDs) do not significantly affect Movement Time (MT). On the contrary, Algo 1,2 and 3 (respectively HornSchunck, LucasKanade for up-down movements and blob tracking for left-right ones), levels of difficulty (IDs) significantly affect Movement Time (MT) (see Figure 3). However, differences are more often significant when ordinal distance between Index of Difficulty is greater than 1 (e.g., Movement Time values (MT) in ID=4 are significantly only higher than MT values in ID=1 and ID=2 cases).

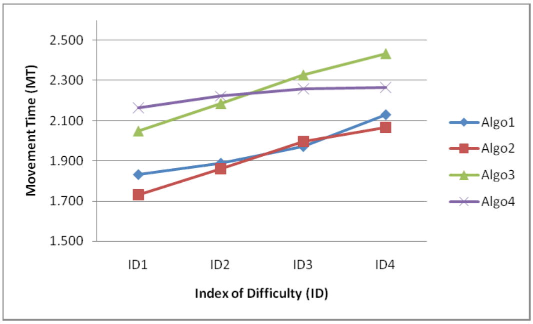

Fig. 4. Interaction plot of Algorithm (Algo) by Index of Difficulty (ID) effects. Algo1 and Algo2 respectively refer to the HornSchunck, LucasKanade solutions for monitoring up-down movements from the face camera; Algo3 and Algo4 respectively refer to blob tracking and Lucas-Kanade solutions for monitoring left-right movement from the top and face cameras.

Accuracy Measures. Following [8], a set of new accuracy measures was developed to supplement traditional ones, such as movement time and error rate, in assessing the Virtual Binocular Interface performances. Current features include directness and geometric entropy indexes to investigate the spread of the trajectories.

- *Directness index* (DI) is computed as the ratio between the length of the straight line connecting the first and last point of a trajectory (in this case, the line between the two targets) and the sum of the lengths of each segment composing the trajectory. It is inspired by the Space dimension of Laban's Effort Theory [2].
- *Geometry Entropy Index* (GEI) is computed by taking the natural logarithm of twice the length of the trajectory (LP) divided by the perimeter of the convex hull around that path:

$$GEI = ln(\frac{2 * LP}{c}) \tag{1}$$

where LP is the path length and c is the perimeter of the convex hull around LP. It is inspired by [10].

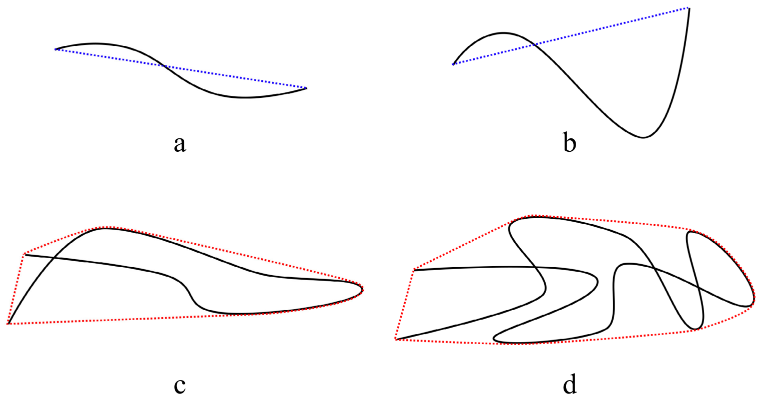

Fig. 5. Four trajectories (continuous black lines) exhibiting, respectively: (a) high DI, (b) low DI, (c) low GEI and (d) high GEI. The blue dotted lines represent the shortest paths between the starting and ending point of trajectories a and b. The red dotted lines represent the convex hulls of trajectories c and d.

In order to assess the validity of these new features, we replicated the statistical analysis conducted on Movement Time. A mixed three-way ANCOVA was performed, first on the Geometry Entropy Index (GEI) and secondly, on the Directness Index (DI) of participants with algorithm (Algo), Index of Difficulty (ID) and Orientation as factors and error rate as covariate.

In the Geometry Entropy Index case (GEI), a main significant effect was found for Algo ($F_{2,2528} = 55.061, p < 0.001$). For the Directness index (DI), main significant effects were found for ID ($F_{3,2528} = 3.866, p < 0.05$), Algo ($F_{2,2528} = 10.055, p < 0.001$) and ID x Algo interaction ($F_{6,2528} = 2.371, p < 0.05$). Post-hoc comparisons revealed that the directness index values (DI) augment with the level of difficulties, and at different intensity according to the considered algorithm. In particular, Algo 1 (Horn-Schunck) showed the highest progression in this respect.

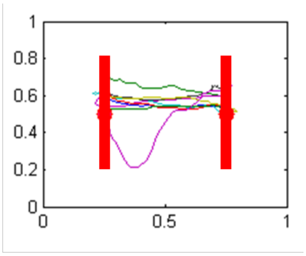

Fig. 6. Sample trajectories from a participant. The participants were instructed to select the red-highlighted target as quickly and accurately as possible. One block consisted of going back and forth between the two targets 10 times.

Questionnaire. The device assessment questionnaire consisted of 12 questions conforming to ISO 9241-9 [12,3]. The questions pertained to the Virtual Binocular. Each response was rated on a seven-point scale, with 7 as the most favorable response and 1 the least favorable response. Results (means and confidence intervals) are shown in Figure 7.

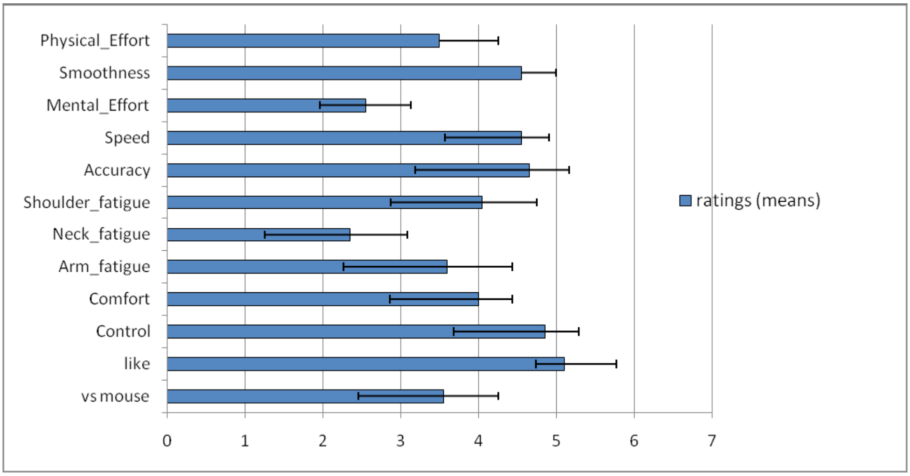

Fig. 7. Participants' ratings of the 12 items questionnaire

These new ratings confirmed that the participants enjoyed using the Virtual Binocular as pointing device (see like item, mean=5.2). The relatively high values obtained for smoothness, speed and accuracy and control (mean respectively of 4.5, 4.5, 4.6 and 4.9) may also suggest that the system efficiently support participants' pointing strategies. A repeated measure ANOVA showed that differences in ratings were significant ($F_{11,209} = 9.92, p < .001$). Bonferroni-corrected post-hoc comparisons revealed that ratings for mental effort and neck fatigue were significantly lower than smoothness, speed, accuracy, control and like items . These results may ensure that the high performance of the system are enjoyed as they can be achieved through intuitive and natural gestural behavior. A confirmation on the good acceptance of the interface is the very high rating of the installation Viaggiatori di Sguardo, where we collected more than 1000 informal reports from visitors, resulting in about 98% of "high satisfaction" and "enjoyment" in the use of the interface in the context of "journey"/"exploration" of the audiovisual content of monumental buildings.

3.4 Discussion

The statistically significant differences between MT with respect to ID values and the observed linear trend may indicate that the modeling of our data can result from the application of Fitts law. In pointing tasks (i.e. for rapid aim movement), the values of IDs is said to predict movement time in a linear way [4]. However, the level of difficulty as assessed by the ID values can be moderated or increased by the technique employed for monitoring the participants movement (see the Lukas Kanade solution for left-right movement tracking). Additional elements may be considered to reach a better understanding of the Virtual Binocular performance. In this perspective, the information obtained through the application of the new accuracy measures (Geometry Entropy and Directness Indexes) confirmed and extend the outputs of the Movement Time analysis. The effect of ID x Algorithm interaction on the participants' performance is for example confirmed. These results also help considering the way in which each ID and each algorithm solution affect the participant movement. According to the algorithm envisaged solution (for example Algo1, Horn-Schunck), a high level of difficulty can foster the participant in reducing the spread of the cursor trajectory. This qualitative information about the participant reactions help characterizing the usability of the Virtual Binocular in an objective manner. Other aspects of movement may relate to more high-level expressive dimension that may indirectly affect the performance, or at least the feeling of the participant in using the Virtual Binocular.

4 Conclusion

This paper presented an evaluation of the novel Virtual Binocular Interface. The binocular is one of the emerging examples of new active experience paradigms developed at our research centre. A sensitive environment equipped with non-invasive tracking technology and following interaction design principles enables a seamless interaction with the content through natural gesture.

We proposed a set of evaluation tools including qualitative analysis of motor performance. Results showed that the Geometry Entropy Index may be particularly suitable to evaluate the usability of an interface that support active embodied exploration in pointing tasks.

Future work includes enhancements to the interface, testing of further new modalities, metaphors, and paradigm of interaction, for increasing the degrees of freedom and possibilities of users in controlling cultural audiovisual content.

Acknowledgments. This work is partially supported by the EU FP7 ICT Project I-SEARCH (A unIfied framework for multimodal contect Search; n°248296; DG INFSO Networked Media Unit; 2010-2012).

References

1. Camurri, A., Canepa, C., Coletta, P., Cavallero, F., Ghisio, S., Glowinski, D., Volpe, G.: Active experience of audiovisual cultural content: the virtual binocular interface. In: ACM Workshop eHeritage 2010, Firenze, Italy (2010)
2. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 460–467. Springer, Heidelberg (2004)
3. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit. ACM (1999)
4. Drewes, H.: Only one fitts' law formula please! In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2010, pp. 2813–2822. ACM, New York (2010)
5. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
6. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17(1-3), 185–203 (1981)
7. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence, vol. 3, pp. 674–679. Citeseer (1981)
8. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 9–16. ACM (2001)
9. Bianchi-Berthouze, N., Mussio, P.: Context and emotion aware visual interaction. Journal of Visual Languages and Computing 16, 383–385 (2005)
10. Cordier, P., Mendes France, M., Pailhous, J., Bolon, P.: Entropy as a global variable of the learning process. Human Movement Science 13, 745–763 (1994)
11. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. International Journal of Human-Computer Studies 61(6), 751–789 (2004)
12. ISO International Organization for Standardization. 9241 9 ergonomic requirements for office work with visual display terminals (vdts)requirements for nonkeyboard input devices (2000)