ORIGINAL ARTICLE

Exploring individual user differences in the 2D/3D interaction with medical image data

**Elena Zudilova-Seinstra · Boris van Schooten ·
Avan Suinesiaputra · Rob van der Geest ·
Betsy van Dijk · Johan Reiber · Peter Sloot**

Abstract User-centered design is often performed without regard to individual user differences. In this paper, we report results of an empirical study aimed to evaluate whether computer experience and demographic user characteristics would have an effect on the way people interact with the visualized medical data in a 3D virtual environment using 2D and 3D input devices. We analyzed the interaction through performance data, questionnaires and observations. The results suggest that differences in gender, age and game experience have an effect on people's behavior and task performance, as well as on subjective user preferences.

1 Introduction

High-end clinical workstations may vary from non-immersive desktop systems to semi- and fully immersive virtual reality (VR) environments (Brooks 1999). However,

E. Zudilova-Seinstra (✉) · P. Sloot
Computational Science, Informatics Institute,
Faculty of Science, University of Amsterdam,
Amsterdam, The Netherlands
e-mail: E.V.Zudilova-Seinstra@uva.nl

B. van Schooten · B. van Dijk
Human-Media Interaction, Department of Electrical
Engineering, Mathematics and Computer Science,
University of Twente, Enschede, The Netherlands

A. Suinesiaputra · R. van der Geest · J. Reiber
Division of Image Processing, Department of Radiology,
Leiden University Medical Center, Leiden, The Netherlands

broad exploration of VR-based medical applications is hampered today by various usability and user-acceptance problems. These arise not only from uncomfortable user interfaces, but also from input/output devices chosen incorrectly for the deployment of an interactive medical environment.

Although medical systems have been among the targeted application areas of VR for years (Jin et al. 2005; Hoffman et al. 2001; Gabbard et al. 1999), VR is hardly used today in the real-life clinical environment. Also, to our knowledge, available literature about the usage of virtual environments within the medical context does not provide much information on the problems and choices encountered when developing VR systems intended for such a specific context, especially with regard to optimal input/output devices.

Very often we do not take into account the fact that clinicians are mostly inexperienced computer users, and therefore they need intuitive interaction support and relevant feedback adapted to their knowledge and everyday skills (Sloot 2000). To provide clinicians with an intuitive environment to solve a target class of problems, a medical application has to be built in such a way that the user can exploit modern technologies without specialized knowledge of underlying hardware and software. Unfortunately, in reality the situation is far from ideal.

Not only 3D user interfaces are generally unfamiliar to medical specialists but also using them brings along new issues that do not come into play when dealing with traditional 2D desktop applications. A complete analysis of a VR-based medical application needs to take into account how the interaction techniques and devices being offered allow the clinician to map his/her high-level objectives and tasks into specific actions that can be interpreted and executed by the system.

To address this research concern, we developed an experimental multimodal visualization framework that supports input and display devices of both VR and desktop systems. It includes the 2D/3D switchable Sharp LL-151-3D auto-stereoscopic monitor and the 2D/3D Essential Reality P5 glove. These devices allow semi-immersive virtual and non-immersive desktop realities to be alternated in a sequential manner.

The paper reports the current implementation status and presents an experimental study conducted to evaluate whether individual user differences (i.e., gender, age, computer experience) have an effect on the way people interact with 3D medical image data while performing interactive steering tasks. Semi-automated medical segmentation served as the context for this research. We compared the virtual P5 glove in a 2D/3D mode and the 2D Logitech PC mouse. Our design was repeated measures within-subjects for input method/device and task complexity. We report our main findings suggesting criteria for applying 2D/3D interaction to a medical exploration environment.

2 Related work

Evaluation has often been the missing component in the field of 3D interaction and visualization (Bowman et al. 2005). For years, researchers focused on the development of new interaction devices, techniques and metaphors for exploring 3D spaces without taking time to assess how good their designs are in comparison to alternative solutions (Johnson 2004).

Prior research has shown that the efficient use of 2D graphical user interfaces strongly depends on human abilities. One of the primary user characteristics that interface designers adapt to is the level of experience or the expert-versus-novice paradigm. Eberts (1994) reports that experts and novices have diverse capabilities and requirements that may not be compatible. Experience level influences the skills of the user, the abilities that predict performance and the manner in which users understand and organize task information (Dix et al. 1993; Egan 1988).

Another adaptive approach addresses the plasticity of human cognitive abilities (Stanney et al. 1998). Several studies suggest that technical aptitudes (e.g., spatial visualization, orientation, memory, etc.) are significant in predicting HCI performance. Leitheiser and Munro (1995) and Vicente et al. (1987) concur that measures of spatial abilities predict performance in a variety of file management tasks, while experience alone does not influence task performance. Gagnon (1985) reported the surprising result that computer game scores were not correlated with hand–eye coordination but were correlated with scores on a spatial memory test. Egan and Gomez (1985) found that measures of spatial memory and age provided the best predictors of how well participants learned to use a text editor.

Spatial abilities as a component of human intelligence have been considered by cognitive psychologists for many years. Consequently, plenty of related studies have been performed. Some well agreed upon findings are that there are considerable differences in spatial abilities among the general population. Velez et al. (2005) report that males on average score better on standard paper tests of spatial abilities, while Salthouse et al. (1990) argue that increased age is associated with lower levels of performance on spatial visualization tests for both unselected adults and adults with extensive spatial visualization experience. According to Lohman (1996), spatial abilities can be improved via training and experience, e.g., playing action video games helps to reduce gender differences in spatial cognition (Feng et al. 2007).

Virtual environments have often been used as a means to study human spatial behavior. Related literature reports on notable individual differences in spatial behavior attributable to computer experience (Wingrave et al. 2005), gender (Larson et al. 1999; Waller 2000) and spatial abilities (Luursema et al. 2008; Rizzo et al. 2000). Relevant research includes comparing spatial information transfer of virtual environments to the real world (Waller et al. 2001) and real world studies such as selecting objects with a laser pointer (Myers et al. 2002). Results show that users are able to exploit spatial abilities and to transfer organizational knowledge to a 3D virtual environment. Furthermore, virtual environments have been used to assess and to treat balance (Jacobson et al. 2001) and psychological disorders (Hodges et al. 2001; Botella et al. 1998), as well as to improve spatial rotation among deaf and hard-of-hearing children (Passig and Eden 2001).

A variety of input devices like data gloves, joysticks and hand-held wands allow the user to navigate through a virtual environment and to interact with virtual objects. Input devices can be characterized by their degrees of freedom (DOF), which describe the possible interaction space (He and Kaufman 1993). 2D input devices (e.g., mouse, joystick, etc.) are bound to the (x, y) plane and have only 2DOF available for interaction. 3D input devices (e.g., space mouse, data glove, phantom, wand, etc.) have 6DOF describing translation of the device along any of three perpendicular axes (x, y, z) and rotation of the device around any of these axes.

There have been several experiments performed to compare 2D and 3D input, which showed that 2D and 3D input devices have their advantages and disadvantages in the sense that some are better suited for certain tasks than others. Most of the studies, however, have been performed across basic manipulation, docking or navigation tasks

(Martens et al. 2007; Bowman et al. 1999; Roessler and Grantz 1998; Hinckley et al. 1997), while realistic medical tasks have been very rarely considered (Krüger et al. 2007; Bornik et al. 2006).

In the medical context, neither usability, nor human-related factors have been sufficiently addressed yet with regard to the choice of input devices. Information on which interaction technique and device are most suitable for a specific medical exploration task is yet too scarce to make an informed choice. Also, it remains unclear whether individual user differences have an effect on the 2D/3D interaction with the visualized medical data, as well as on subjective user preferences for available input methods.

3 A multimodal visualization framework for medical image analysis

Virtual and desktop realities are alternative solutions that allow users to manipulate and navigate through visualized datasets. Even though both virtual and desktop systems are viable alternatives for the image-based exploration, none of them is able to provide optimal means for analyzing medical data. In our prior research (Zudilova and Sloot 2005), we discovered that for the medical exploration tasks, where the insight view or collaboration between clinicians is important, VR would be the best choice. But when performance and accuracy are vital, the medical application running on a desktop system is usually preferable.

Having this in mind, we developed a multimodal visualization framework that supports features of both desktop and VR systems. As can be seen in Fig. 1, right, the framework does not require much space. It is portable and relatively cheap, which makes it a valuable option for hospitals, as they usually do not have sufficient space and budget available for more complex VR configurations (Cramer et al. 2004).

The framework includes a 15 in. auto-stereoscopic monitor Sharp LL-151-3D providing a view on a virtual environment (http://www.vrealities.com/sharpll1513d.html). The Sharp's TFT 3D LCD Technology makes the image on the screen appear in 3D without the need for the user to wear special glasses. The display can be set to monoscopic or stereoscopic viewing modes electronically, offering a single display for both VR-based visualization and normal 2D work.

The handling of objects in a 3D virtual environment typically involves manipulation and system control, which often supports manipulation itself. The multimodal visualization framework uses keyboard input for the system control and allows mixing of the glove and mouse input for direct manipulations. We chose a P5 Glove Controller from
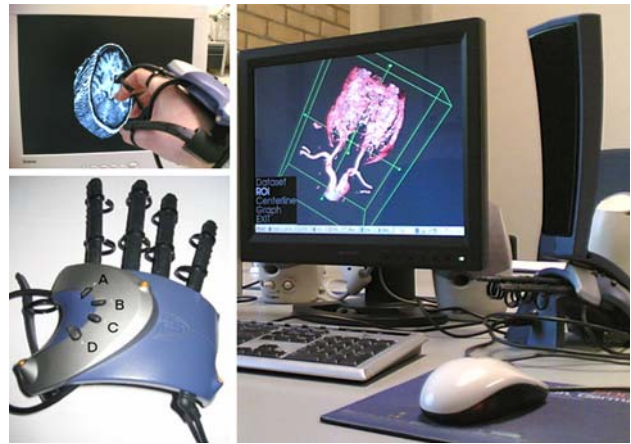

Fig. 1 A multimodal visualization framework for medical image analysis (*right*) and the virtual P5 glove (*left*)

Essential Reality (http://www.vrealities.com/P5.html) because it is a switchable device that can be used to control both 2D and 3D input.

The virtual P5 glove features five bend sensors to track bending of the user's fingers and an infrared-based optical tracking system allowing computation of the glove position and orientation with the frequency of 60 times/s. The glove consists of a base station housing infrared receptors enabling spatial tracking. The glove itself is connected to the base station with a cable and consists of a plastic housing that is strapped to the back of the user's hand, with five bendable strips connected to the fingers to determine the bend of each individual finger. The glove has 2.4 mm resolution and 9.7 mm accuracy for position and 1° resolution and accuracy for orientation measurements. Also, it provides five single joint independent finger measurements with 0.5° resolution. The measurement of a finger bend returns an integer value in the range [0, 63]. These values can be personalized in a quick calibration phase, such that they are converted to the actual finger bending of each user. Furthermore, on top of the housing there are four buttons that can be used to provide additional functionality (see Fig. 1, left).

The main disadvantage of the virtual P5 glove is a relatively small range from the receptor (1.5 m) that allows accurate tracking of position and orientation and the tracking of a single joint finger bending. However, since the user usually sits about 40 cm away from the computer screen, this disadvantage is easily dissolved by putting the infrared receptor next to the screen. Also, during our work with the glove, we learned that the spatial tracking data were not always reliable. To ascertain sufficiently reliable values, additional filtering mechanisms were developed, including the dynamic averaging procedure based on the rate of changes in motion and rotation data.

The semi-automated segmentation of the medical images of the patients suffering from atherosclerosis serves as the context for our research. Image segmentation techniques are essential to provide objective quantitative data characterizing a vascular abnormality. In many pathologies, a limited number of quantitative parameters describe the relevant clinical findings from the imaging study. Like any image processing system, automated segmentation algorithms can produce mistakes, e.g., when the contrast level or the amount of noise differs from those specified or when the bifurcations or closely located vessels are misinterpreted by the algorithm and as such affect the segmentation result (Adame et al. 2004). To obtain correct measurements, manual adjustments or overwrites to automated segmentation results are frequently needed in routine clinical practice. Often manual editing is a time-consuming and tedious procedure and affects the otherwise objective measurements. To overcome these problems, semi-automated image processing techniques need to be integrated in the data exploration process such that image segmentation, visualization and user steering become a unified process.

The multimodal visualization framework developed in this project is built on the principle that the user (clinician) will be able to alternate desktop and virtual realities while performing interactive steering tasks related to medical segmentation, e.g., selection of the region of interest, interactive placement of seed points, labeling, centerline correction, contour editing, etc.

4 Method

All users of an interface bring their preferences, aptitudes (physical, perceptual and cognitive) and prior experiences in the world. These attributes can be considered as distinct from, but interacting with, the user's knowledge and skills that result from direct experience, practice, feedback and training on an interface (Wingrave et al. 2005).

The present study examines whether individual user differences influence task performance and subjective preferences for 2D/3D input methods, applied to manipulate the visualized medical data. By having participants surveyed for demographic information (age, gender, education, etc.), as well as for computer experience (computer use, game experience, experience with 3D graphics, etc.) and physical characteristics (hand dominance, acuity of vision, etc.) that have potential relations to skills needed to perform specific interactive steering tasks, we can begin to uncover predictors of performance and highlight user attributes that may influence the choice of input methods/devices and design of a virtual medical environment in general.

4.1 Tasks

To perform the study, we chose two interactive steering tasks related to medical segmentation: selection of the region of interest (selection task) and correction of the automatically generated centerline (positioning task). These tasks were selected for two reasons. Both tasks are frequently performed and are crucial for the successful completion of the segmentation process. Also, for these tasks, objectives can be precisely defined and potentially confounding factors can be controlled.

We compared the virtual P5 glove in a 2D/3D mode and the 2D Logitech PC mouse. In the tasks, participants had to manipulate so-called 3D widgets. Simply speaking, a widget is an object in a scene that responds to user events (e.g., mouse clicks) and data changes by corresponding changes in its appearance or behavior (Conner et al. 1992). 3D widgets make the user interaction with 3D objects more intuitive by providing fast semantic feedback.

In the selection task, participants had to select the region of interest by manipulating a 3D box widget. In the positioning task, they had to adjust the position of a centerline by manipulating a 3D spline widget. 3D box and spline widgets are shown in Fig. 2 and described in more detail in the next section.

4.2 The widget interface

The custom experimental environment was developed using the kitware visualization toolkit (VTK), where different types of widgets require their own way of interaction (Schroeder et al. 2002).

Represented by an arbitrarily oriented hexahedron with orthogonal faces, a 3D box widget defines a region of interest (Fig. 2, left). It has seven handles that can be manipulated. The first six correspond to the six faces and can be used for the face-based scaling of the widget. By grabbing these six face handles, faces can be moved in the direction of one of three axes (x, y or z) depending on the handle position.
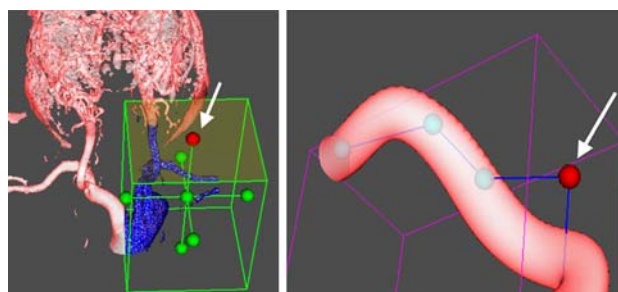

Fig. 2 A 3D box widget applied to the selection task (*left*), a 3D spline widget applied to the positioning task (*right*)

The seventh handle is in the center of the hexahedron. By grabbing the central handle, the entire hexahedron can be translated in 3D space. With 2DOF input, the positioning of the hexahedron in 3D space requires two sequential actions: translation of the central handle in the (x, y) plane combined with scene rotation, performed in the direction defined from the center of the viewport toward the cursor position. With 6DOF input, the positioning of the hexahedron in 3D space can be achieved via one atomic action with a 3D input device.

In addition, all faces of the hexahedron can be manipulated. These allow the face-based rotation of the hexahedron. With 2DOF input, face-based rotation is determined by x and y coordinates of the input device, which implies that orientation of the hexahedron can be adjusted, while the position of the central handle remains the same. For instance, starting from the initial condition (Fig. 3a), the user selects the upper face (Fig. 3b) and drags the cursor down with a 2DOF input device causing the hexahedron to be rotated around the x-axis (Fig. 3c). With 6DOF input, face-based rotation is determined by x, y and z coordinates and orientation of the input device and performed in such a way that orientation of a selected widget face is always identical to that of the input device (Fig. 3d).

A 3D spline widget has spherical handles that can be translated to change the shape of the spline (Fig. 2, right). With 2DOF input, each handle can be translated only within the (x, y) plane. With 6DOF input, handles can be freely translated and oriented in 3D space. By picking on a line segment, forming the spline, the complete spline can be translated. The translation of the spline in 3D space is performed in a similar way as the 3D box widget translation explained earlier.

Visualization toolkit allows both widgets to be controlled via a standard PC mouse. By moving the mouse while keeping the left button pressed, widget elements (e.g., handle, line segment, face, etc.) can be manipulated. Scaling is achieved by using the right mouse button "up"

the render window (makes the widget bigger) or "down" the render window (makes the widget smaller).

The VTK C++ hierarchy has been extended with new classes to support the P5 glove 2DOF/6DOF interaction with 3D box and spline widgets and to record the user interaction data. To optimize time required for training subjects, we decided to mostly use buttons (Fig. 1, left) for the widget control. In particular, to select the widget element, the user has to press the button "A" when the cursor reaches the element that has to be selected. In a 3D mode, the widget element can be manipulated by changing position and orientation of the glove. In a 2D mode, the virtual P5 glove functions as the mouse. To deactivate selection, the button "B" should be pressed. Scaling can be achieved by keeping the button "C" pressed and changing the position of the glove, while moving the arm up and down, vertically.

With the mouse, scene rotation occurs continuously as long as the mouse left button is pressed. With a glove, rotation starts when the wearer's index finger is bended to a certain degree. Rotation stops, when the index finger is no longer bended.

5 Experimental setup

Our experiment aimed at quantifying hypotheses formulated based on the argumentation from previous literature. We decided to focus on gender, age and computer experience-related differences for two reasons. These individual user characteristics are easily observed and they are often considered as categorical distinctions for noting differences in people's spatial abilities (Strong and Smith 2001; Hartman et al. 2006).

Before running the experiment, we defined the following four hypotheses:

H1: Due to different spatial abilities of men and women (Luursema et al. 2008; Waller 2000), it is expected that gender will have an influence on the task completion time
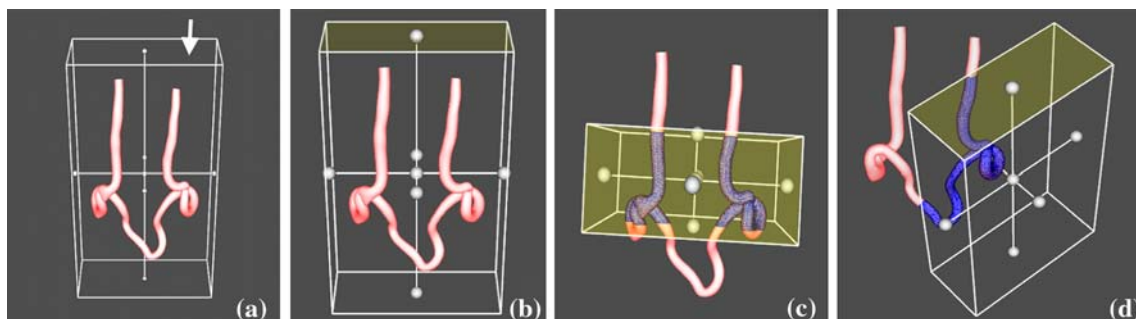

Fig. 3 Illustrations of the face-based rotation technique

for both selection and positioning tasks, as well as on the people's choice for available interaction strategies, where strategies are observed behaviors that certain participants performed to increase their overall performance.

H2: Age is expected to play a role in both time and accuracy. Due to the age-related differences in mental rotation (Berg et al. 1982), we expect that younger subjects will be able to benefit more from the usage of the virtual P5 glove for the interaction with a 3D virtual environment than older subjects. This may also influence user preferences for available input methods/devices.

H3: Computer experience (i.e., game experience, computer and graphics usage) is expected to have an effect on task performance, as well as on subjective user preferences (Wingrave et al. 2005).

H4: Due to different demands (i.e., visual display, complexity, etc.) imposed upon the users by selection and positioning tasks, the influence of individual user differences will vary depending on the task being performed (Karwowski 2006).

5.1 Experimental conditions

Our design was 3 × 3 repeated measures within-subjects for input method/device and task complexity. The virtual P5 glove in a 3D mode (3D glove) was tested against the Logitech PC mouse and the virtual P5 glove in a 2D mode (2D glove). We experimented with one 6DOF (3D glove) and two 2DOF (mouse and 2D glove) input methods. The 2D glove condition was included to ensure that our results would not be biased due to the prior intensive mouse experience of participants and resolution differences of devices. The order of input methods was counterbalanced to prevent carry-over effects (e.g., learning or fatigue).

In our study, we applied the evaluation methodology introduced by Moise et al. (2005). According to Moise et al. (2005), it is possible to test the radiology workstation interaction features using look-alike radiological tasks and inexperienced laypersons, and that the results transfer to radiologists performing the same tasks. We adjusted the custom experimental environment in such a way, that selection and positioning tasks could be easily interpreted and performed by people without medical background and ran a small pilot study to make sure that our experiment was indeed suitable for laypersons.

In the selection task, participants were asked to select the specified region of interest. To achieve this, they had to manipulate a 3D box widget, initially positioned in such a way that all vessel structures displayed on the screen were covered by the widget. We introduced three complexity levels (low, medium, high) for each task. The complexity of the selection task was defined by the number and density

of vessels, from which participants had to choose the correct vessel segment (see Fig. 4):

- Level 1 (low)—one vessel;
- Level 2 (medium)—two closely located vessels;
- Level 3 (high)—three vessels, where two vessels are closely located.

In the positioning task, participants were asked to adjust the position of a centerline represented by a 3D spline widget in such a way that all spline handles would be located inside the vessel segment. Initially, the 3D spline widget was located such that only the first and last handles were positioned correctly. To allow participants to easily notice positioning problems, we used occlusion cues. The complexity of the task was defined by length and curvature of the vessel segment and the number of handles, which positions had to be adjusted (see Fig. 5):

- Level 1 (low)—a five-handle spline widget has to be positioned inside a small vessel branch;
- Level 2 (medium)—a nine-handle spline widget has to be positioned inside a mid-size highly curved vessel branch;
- Level 3 (high)—a 12-handle spline widget has to be positioned inside a large-size vessel branch curved in the middle.
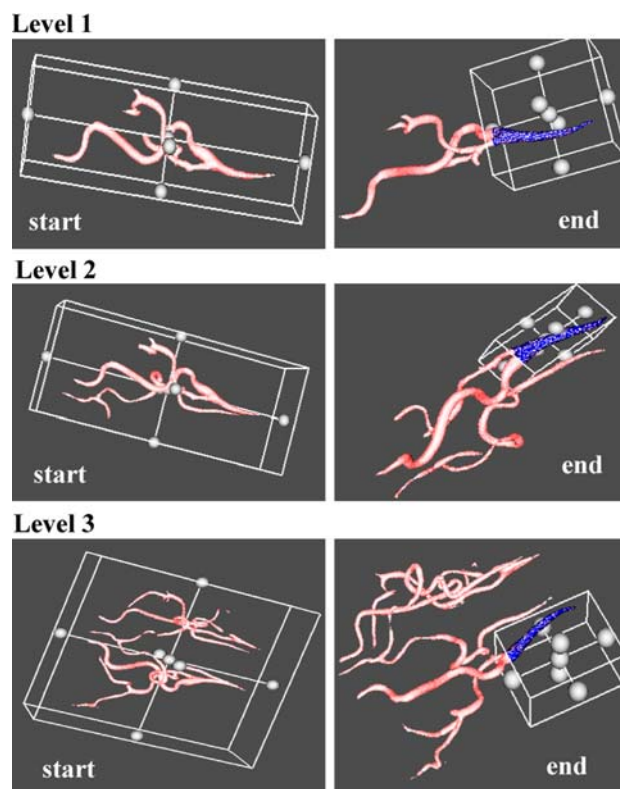

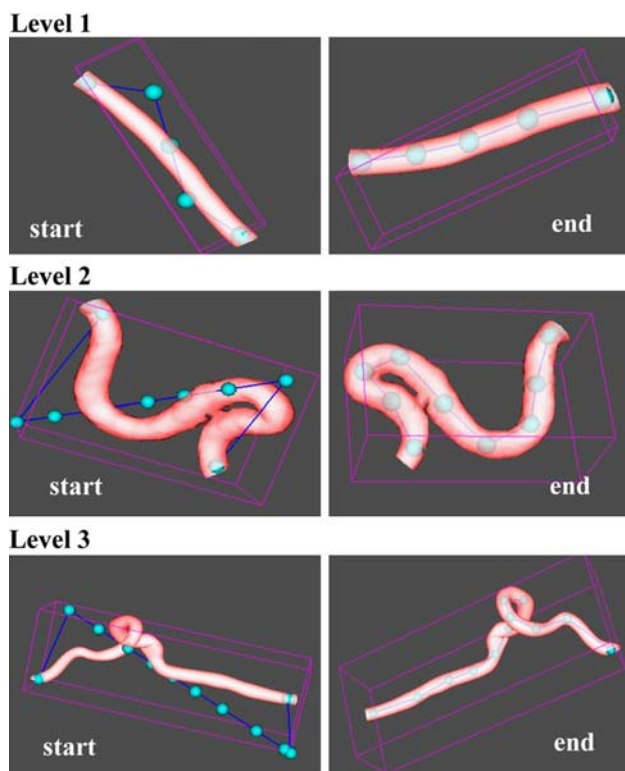**Fig. 4** Illustrations of the selection task in the 3D widget experiment

Fig. 5 Illustrations of the positioning task in the 3D widget experiment

5.2 Procedure

A total of 30 volunteers (13 female, 17 male) were recruited from different departments of the Informatics Institute. All participants performed both selection and positioning tasks, which were assigned in a random order. Participants had varied levels of computer and graphics usage, as well as of game experience. None reported previous experience with 3D data glove devices. 29 participants were right-handed. One was ambidextrous.

The experimental sessions consisted of four trial series and lasted approximately 45 min. Participants first reviewed instructions and completed a prior-trial on-line questionnaire. The following information has been collected for each participant: name, age, gender, background, hand dominance, acuity of vision, computer use, gaming experience, experience with graphics and interaction devices.

Then participants received a short demonstration regarding tasks and the interface. Before each trial series, participants completed a training session to get familiar with the task and the input method. For each condition, trials were assigned in ascending order of task complexity to provide optimum conditions for the task-related skill development and efficient scheduling of task performance components (Robinson 2001).

Participants were instructed that time and accuracy were of equal importance and provided with the indication of accuracy for both tasks. However, no indication was given what would be fast enough. The precise definition of performance was left to their own judgment. When satisfied with the result, participants selected the next trial (complexity level) from the system menu.

Dependent variables were the task completion time, accuracy and subjective ratings of the ease-of-use and preference. We defined the task completion time as the duration between the moment when the current trial was loaded and the moment when the next trial was selected from the menu. Accuracy of the selection task was measured via a surface-based comparison of the selected region of interest and the ideal result with allowed precision of 5%. To measure accuracy of the positioning task, the number of line segments were counted that either had one intersection with or were positioned completely outside the vessel segment.

In a post-trial questionnaire, participants rated 2D/3D input methods available for selection and positioning tasks and indicated their preferences. Subjective ratings were administered using a four point scale and open-ended questions. We used the SurveyMonkey.com (http://www.surveymonkey.com) online tool to develop questionnaires and to collect responses.

In addition, a few specific measurements were taken per each condition, including the scene rotation time and the total interaction time. The interaction time is the actual time spent on direct manipulation. Furthermore, we measured the face-based scaling time, the face-based rotation time and the hexahedron translation time for the selection task. For the positioning task, we measured the handle translation time and the time spent on spline translation.

6 Results

In this section, we discuss our results suggesting that differences in gender, age and game experience have an effect on user behavior and performance, as well as on subjective user preferences. In the sample of 30 study participants, there were no correlations among these three independent variables found.

For testing significance, we used analysis of variance (ANOVA) and post hoc Tukey HSD (honestly significant difference) tests. We applied Tukey HSD tests for pairwise comparisons because the Tukey HSD test is more sensitive when making large number of comparisons than other commonly used post hoc tests, i.e., Bonferroni t tests (Plichta and Garzon 2008). Timing data were transformed using a natural logarithm to improve the fit to a normal curve and then analyzed using repeated measures factorial

3 × 3 ANOVA. Repeated measures factorial ANOVA was applied because each subject was tested in all conditions. When Mauchly's test of sphericity indicated it was necessary, we used the Huynh–Feldt correction. For rating scale data, the non-parametric Kruskal–Wallis ANOVA by Ranks was applied.

6.1 Gender issues

The effect of gender turned out to be slightly less than we expected. In particular, we do not have enough statistical evidence to support the first part of hypothesis H1, stating that the task completion time was affected by gender differences. However, our results suggest that gender had a significant effect on the total interaction time measured for both selection and positioning tasks (see Fig. 6).

ANOVA revealed that female subjects spent significantly more time on the actual interaction with a virtual environment while performing selection tasks than male subjects, $F(1, 29) = 7.4$, $p = 0.011$. The average interaction time was 24.8 s for women and 21.32 s for men. There was also the main effect of task complexity on the interaction time, $F(1.85, 51.81) = 91.11$, $p < 0.001$, indicating that more complex selection tasks were generally much longer and as such required more time to be spent on direct manipulation by both men and women. Post hoc Tukey HSD tests (at $p \leq 0.05$) were conducted to examine further the effect of task complexity on the gender-related differences in the interaction time. A significant difference between men and women in the average time spent on interaction was found only for the most complex trial of Level 3.

In the positioning task, the average interaction time was 65.07 s for women and 46.74 s for men. ANOVA found the significant interaction effect between gender and input method [$F(2, 56) = 3.3$, $p = 0.044$], suggesting that the difference in the total interaction time between female and male subjects was greater with the 2D glove than with the 3D glove and with the 3D glove than with the mouse. Also, we found the main effect of task complexity on the interaction time, $F(1.85, 51.8) = 40.6$, $p < 0.001$, indicating that both men and women spent significantly more time on direct manipulation when performing more complex trials of the positioning task. Tukey HSD tests (at $p \leq 0.05$) revealed that the average time spent by male subjects on direct manipulation while performing more complex trials of Level 2 and Level 3 was significantly lower compared to female subjects.

Hence, women spent more time on the actual interaction with a virtual environment than men while performing both selection and positioning tasks. These can be due to different interaction strategies chosen by male and female subjects.

As mentioned in Sect. 4.2, participants had a choice of three manipulation techniques that could be used (separately or in combination) to perform the selection task: face-based scaling, face-based rotation and hexahedron translation. Although any of these techniques in principle allows the correct result to be achieved, they require from users different skills. For instance, face-based rotation and hexahedron translation require more motor skills than face-based scaling. On the other hand, face-based scaling usually generates more errors and as such requires more precision and decision-making.
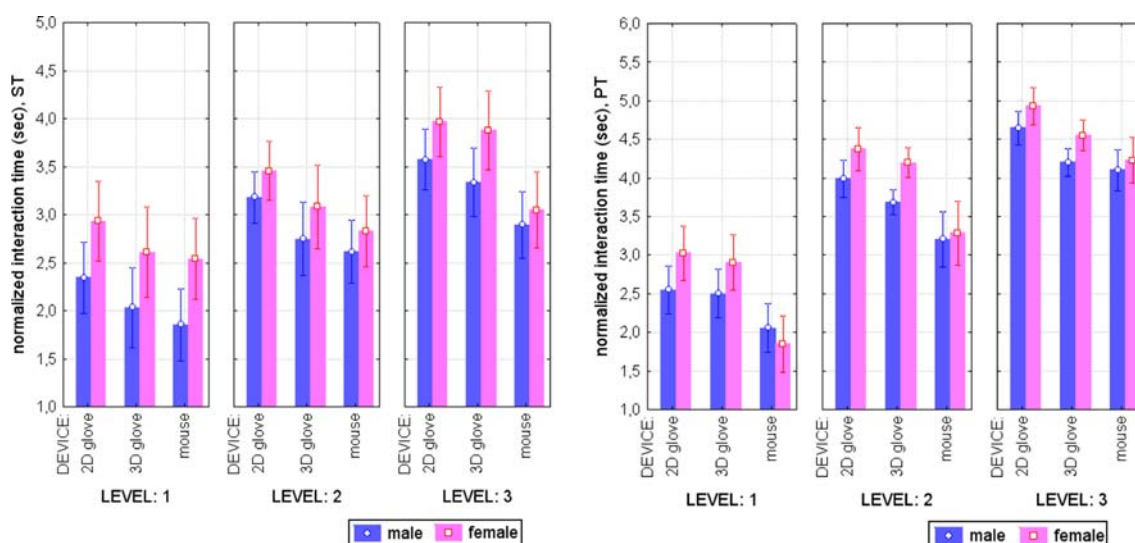

Fig. 6 Average interaction time for the selection (*left*) and positioning (*right*) tasks. Timing data are normalized by log transformation; *vertical bars* denote 0.95 confidence interval

To check whether men and women made different choices for manipulation techniques, we combined timing data obtained for face-based scaling, face-based rotation and hexahedron translation and then analyzed these merged data via $3 \times 3 \times 3$ repeated measures factorial ANOVA (see Fig. 7).

Overall, the face-based scaling technique was frequently applied by men and women to perform selection tasks. However, while male subjects mostly used face-based scaling, female subjects actively combined all three manipulation techniques, especially when performing more complex trials. ANOVA found the significant interaction effect between gender and the type of manipulation [$F(2, 38) = 7.5$, $p = 0.02$], suggesting that face-based scaling was indeed used significantly more by men than by women, while with two other manipulation techniques it was other way around. These eventually resulted in a significant difference in the average interaction time.

Tukey HSD tests (at $p \leq 0.05$) revealed that participants mostly used face-based rotation to perform complex selection tasks with the 2D/3D glove, while with the mouse this technique was applied significantly less. This can be partially explained by resolution/accuracy differences of input devices. The Logitech PC mouse maps the position of the hand, while the virtual P5 glove maps its movement. Consequently, the precise cursor positioning required by the face-based scaling and hexahedron translation techniques becomes more difficult to achieve using the 2D/3D glove, while the least sensitive to the cursor position

technique, face-based rotation, can be always easily performed.

In the positioning task, the difference in the average interaction time was mostly due to the different amount of time spent on scene rotation by male and female subjects (see Fig. 8). We observed during testing sessions that many male subjects spent quite some time on reasoning what the best viewpoint would be to perform the task, while female subjects mostly preferred "multiple probes and trials" approach. Consequently, men (mean 9.75 s) spent less time on scene rotation than women (mean 17.48 s). ANOVA found the main effect of gender on the average scene rotation time, $F(1, 28) = 4.23$, $p = 0.049$. Both men and women spent significantly more time on scene rotation when performing more complex trials, $F(1.92, 53.89) = 87.02$, $p < 0.001$.

There was also the main effect of input method on the scene rotation time, $F(1.95, 54.6) = 8.09$, $p = 0.001$, indicating that significantly less time was spent on rotation with the 3D glove than with the mouse and with the 2D glove. Significant differences in the scene rotation time for women at $p \leq 0.05$ were 6.54 s between the 3D glove and the mouse and 16.27 s between the 2D and the 3D glove. For men, the scene rotation time for all three input methods was not significantly different from each other. These results suggest that female subjects benefited from 3D input more than male subjects, as women spent significantly less time on scene rotation using the 3D glove compared to other input methods.
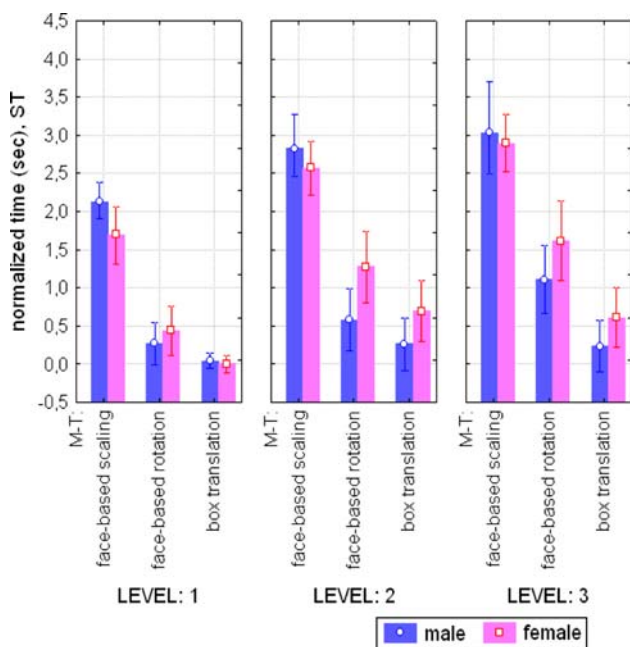

Fig. 7 Average time spent on face-based scaling, face-based rotation and box (*hexahedron*) translation. Timing data are normalized by log transformation; *vertical bars* denote 0.95 confidence interval
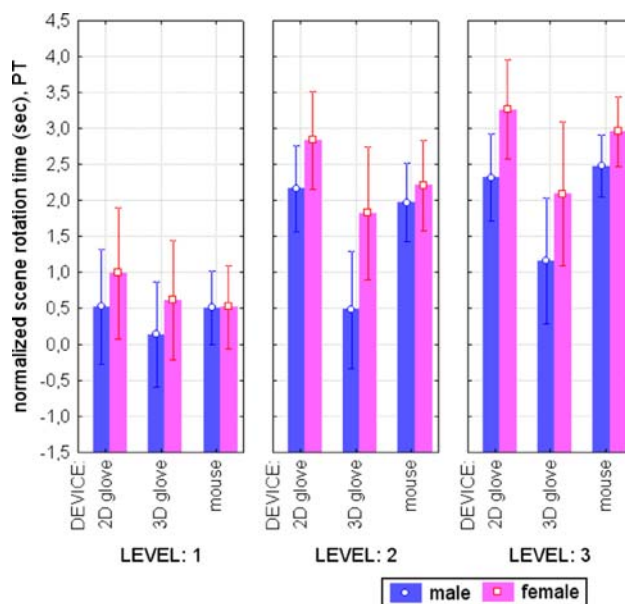

Fig. 8 Average scene rotation time for the positioning task. Timing data are normalized by log transformation; *vertical bars* denote 0.95 confidence interval

6.2 Age issues

The age of participants varied from 19 to 45 years old. Mean age was 28. To perform statistical analysis, we split participants in two age groups: "≤28 years old" (63%) and ">28 years old" (37%). We expected that the age difference would have an effect on task performance (i.e., time and accuracy), as well as on subjective user preferences for available input methods/devices (hypothesis H2).

In general, the experimental data does not contradict our hypothesis H2. However, there was not enough statistical evidence revealed to reason about the effect of age on error data. The rest of hypothesis H2 is well supported by our results.

In the selection task (see Fig. 9, left), the average task completion time for younger subjects (mean 50.19 s) was significantly shorter than for older subjects (mean 84.24 s). ANOVA found a significant main effect for age, $F(1, 28) = 6.65$, $p = 0.015$, and the significant interaction effect between input method and task complexity [$F(3.69, 103.29) = 7.19$, $p < 0.001$], indicating that more complex selection tasks were performed significantly slower with the 2D glove than with the mouse or with the 3D glove by both age groups. The fact that the task completion time with the 2D glove was higher than with any other input method can be explained by the difference in DOF between the 2D and 3D glove, as well as by different prior experiences of subjects with the virtual P5 glove and with the mouse.

A significant difference at $p \leq 0.05$ between younger and older participants in the task completion time was found for the most complex trial of Level 3. Tukey HSD tests also revealed that the trial of Level 3 of the selection task was performed significantly faster by younger subjects with the glove in a 3D mode than with the 2D glove.

In the positioning task (see Fig. 9, right), the average task completion time was much longer for older subjects (mean 173.7 s) than for younger subjects (mean 122.44 s). ANOVA found a significant main effect for age, $F(1, 28) = 5.8$, $p = 0.023$, and the main effect of the trial type, $F(1.93, 54.08) = 52.14$, $p < 0.001$, indicating that more complex positioning tasks generally required more time. Tukey HSD tests (at $p \leq 0.05$) revealed that, for the trials of Levels 2 and 3 of the positioning task, the average completion time was significantly lower for younger people than for older people.

Positioning tasks were performed significantly faster with the mouse than with the 3D glove and significantly faster with the 3D glove than with the 2D glove by both age groups, $F(2, 56) = 53.4$, $p < 0.001$. Post hoc tests (at $p \leq 0.05$) revealed that older subjects were significantly slower with the glove in a 2D mode compared to the glove in a 3D mode and to the mouse. For younger subjects, the only discovered statistically significant difference was between the 2D glove and the mouse.

Due to extensive prior experience of participants with the mouse and lack of experience with the virtual P5 glove, the mouse turned out to be the best device for performing both selection and positioning tasks. However, under the condition of similar user experiences with 2D and 3D devices (2D/3D glove), our results suggest that 3D input is more beneficial for the positioning task than for the selection one. In the positioning task, both age groups performed trials significantly faster with the 3D glove compared to the 2D glove, irrespective of the trial type. In the selection task, 3D input was only beneficial for younger
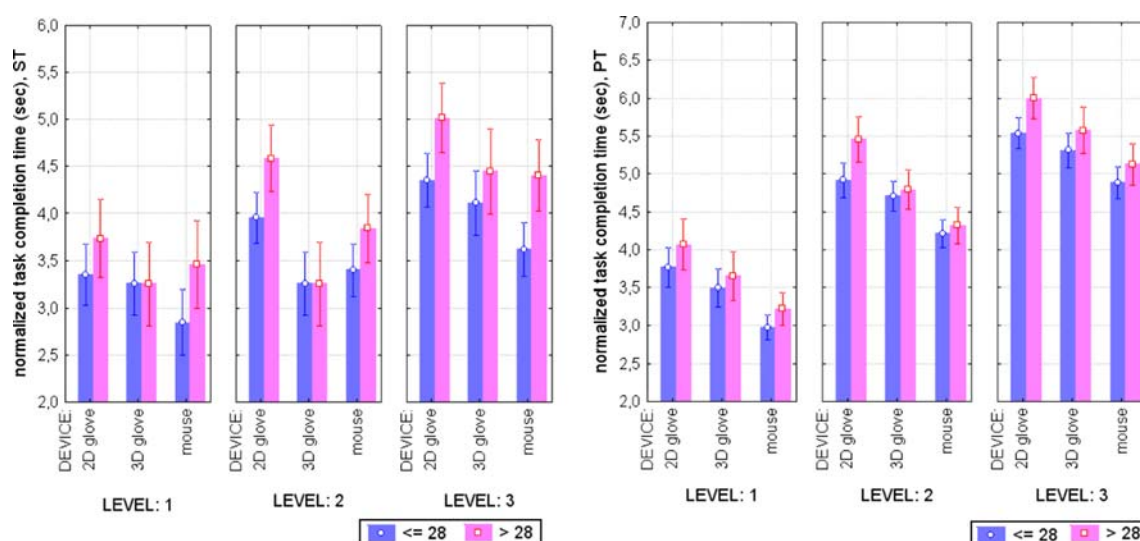

Fig. 9 Average task completion time for the selection (*left*) and positioning (*right*) tasks. Timing data are normalized by log transformation; *vertical bars* denote 0.95 confidence interval

people while they were performing the trial of the highest complexity.

We then analyzed scale data of the ease-of-use ratings of input methods/devices and user preferences for 2D/3D input in general. Although younger and older participants rated input methods differently, there was not enough statistical evidence found to reason about the influence of the user's age on the ease-of-use ratings. With regard to subjective user preferences for 2D/3D input, the Kruskal–Wallis ANOVA by Ranks revealed some significant age-related differences (see Fig. 10).

In the selection task, many young subjects did not have any preference, while all older subjects indicated their preferences for 2D/3D input. These resulted in a main effect for the user's age, $H = 7.19$, $p = 0.007$, suggesting that older people more clearly expressed their preferences than younger ones.

In the positioning task, younger subjects preferred 3D input significantly more than older subjects, and vice versa, older people preferred 2D input significantly more than younger ones. The Kruskal–Wallis ANOVA by Ranks showed a significant age difference in user preferences for 2D ($H = 7.71$, $p = 0.006$) and 3D ($H = 6.8$, $p = 0.009$) input.

These results also support our hypothesis H4 about task-related differences in the influence of individual user characteristics. In the positioning task, younger people were significantly more positive about their experience with the virtual P5 glove in a 3D mode than older people. While in the selection task, age-related differences in user preferences for 2D/3D input were not significantly different. Also, in the selection task, 3D input was more beneficial for younger subjects than for older ones (with respect to the task completion time), which was not the case in the

positioning task, where both age groups benefited from the 3D glove in a similar way.

6.3 Computer experience

The hypothesis about the influence of computer experience (H3) is only partially supported by the experimental data. According to our results, main effects for computer and graphics usage were non-significant. However, there appeared to be some transfer from game experience to subjective ratings of the ease-of-use of input methods/devices as can be seen in Fig. 11.

Game experience was administered using a three point scale, as well as open-ended questions. Most subjects claimed that they had prior game experience: 46% played games occasionally, 30% played games intensively. Only 23% had no prior game experience at all.

The Kruskal–Wallis ANOVA by Ranks showed a significant difference between subjects with different game experiences for the selection task ($H = 6.04$, $p = 0.049$), indicating that the average score of the ease-of-use received from the subjects with no game experience was significantly lower than the average scores from the subjects with occasional game experience and from experienced gamers. On average, the mouse was rated the highest by all three groups, i.e., 1.4–0.9 points higher than the glove in a 3D mode and 1.71–1.34 points higher than the glove in a 2D mode (see Fig. 11, left).

In the positioning task, the Kruskal–Wallis ANOVA by Ranks showed a significant difference in the average ease-of-use ratings of the 3D glove between subjects with different game experiences ($H = 9.77$, $p = 0.008$). Subjects with intensive game experience rated the glove in a 3D mode on average higher than any other input method, while

Fig. 10 Percent of 2D/3D input preferences for the selection and positioning tasks
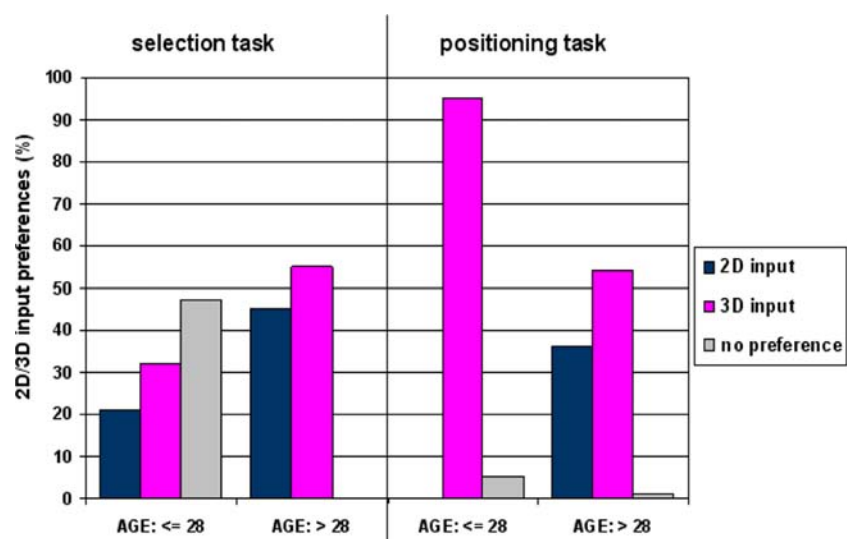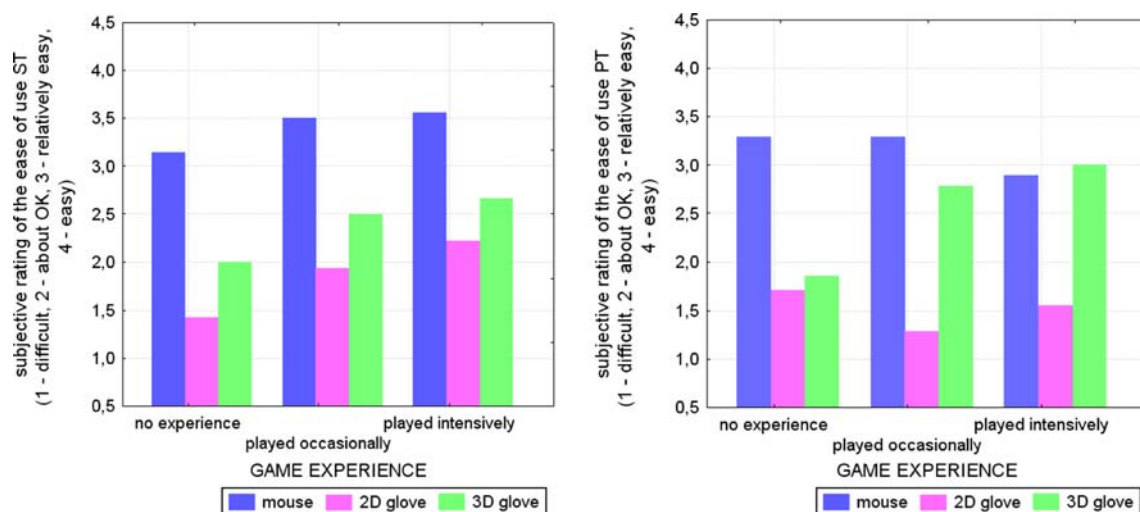
Fig. 11 Average ratings of the ease-of-use of the input devices for the selection (*left*) and positioning (*right*) tasks

less experienced subjects preferred the 2D mouse-based interaction the most (see Fig. 11, right).

As can be seen from above, game experience had different effect on subjective user ratings of input methods/devices depending on the task. In the positioning task, experienced gamers rated the glove in a 3D mode significantly higher than less experienced participants. While in the selection task, the 2D mouse-based interaction was preferred the most, irrespective of differences in game experience between the groups.

Pairwise comparisons of the device condition for the task completion time at $p \leq 0.05$ revealed that people with some occasional game experience performed the positioning task significantly faster using the 3D glove than people without any game experience. While for the selection task, the average task completion time for the glove in a 3D mode was not significantly different from the mouse and from the 2D glove. These imply that subjects with more intensive game experience were able to benefit from the 3D glove much more while performing the positioning task.

Hence, our results suggest that hypothesis H4 is well supported by the experimental data not only with respect to age (see Sect. 6.2) but also with respect to game experience.

7 Discussion and conclusions

In this paper, we investigated the influence of individual user differences (i.e., gender, age, computer experience) on the way people interact with a 3D medical virtual environment while performing interactive steering tasks. The semi-automated segmentation of the patient vascular condition served as the context for this research.

We conducted an empirical study, where participants were asked to perform two tasks: selection of the region of interest (selection task) and correction of the automatically generated centerline (positioning task). Both tasks are part of the semi-automated medical segmentation process and important for its successful completion. We tested the virtual P5 glove in a 2D/3D mode against the Logitech PC mouse.

Our results suggest that gender plays an important role in the user interaction with the visualized medical data. We found the main effect of gender on the average interaction time for both selection and positioning tasks. In particular, female subjects spent significantly more time on the interaction with a virtual environment compared to male subjects. The results indicate that the difference in the average interaction time between men and women was greater as task complexity increased. In the positioning task, the gender-related difference in the average interaction time was greater with the glove in a 2D mode than with the 3D glove. The latter suggests that providing additional DOF for performing positioning tasks may potentially help to reduce differences in the spatial behavior between men and women.

The experimental data showed that differences in the total interaction time were mostly due to different interaction strategies preferred by men and women. In the selection task, women were more inclined to experiment with alternative manipulation techniques than men. In the positioning task, male subjects spent significantly less time on scene rotation than female subjects, which can be explained by the fact that men were more focused on finding the best viewpoint than women (according to our observations).

We also found that the task completion time was significantly affected by age. In particular, younger people

(≤28 years old) were able to perform selection and positioning tasks significantly faster than slightly older people (>28 years old). Our results suggest that 3D input was more beneficial for the positioning task than for the selection task. In the positioning task, both age groups performed trials much faster with the 3D glove than with the 2D glove, irrespective of task complexity. In the selection task, 3D input was mainly beneficial for younger people, when they were performing the most complex trial. Furthermore, in the positioning task, younger people were significantly more positive about their experience with the virtual P5 glove in a 3D mode than older people.

Statistical analysis revealed that game experience had an influence on subjective user ratings of the ease-of-use of input methods/devices. On average, subjects with more intensive game experience rated input devices significantly higher. Game experience had much stronger effect on subjective ratings for the positioning task than for the selection task. In the positioning task, people that played games intensively gave the virtual P5 glove in a 3D mode the highest rates. While in the selection task, the mouse got the highest average rates from all three groups.

Overall, the experimental data suggest that young people and people with prior game experience were able to benefit from the virtual P5 glove in a 3D mode the most and that in general 3D input was more beneficial for the positioning task than for the selection task. Moreover, it appeared that the 3D glove was especially advantageous for female subjects for performing positioning tasks, as they had to spend significantly less time on scene rotation compared to other input methods. As such, we argue that these specific user groups should be provided with a possibility to perform positioning tasks using a 6DOF input device (e.g., P5 glove in a 3D mode).

As for the selection task, it is less clear from the data obtained whether the choice of a certain input method/device can be controlled by the individual user differences explored in this paper. Hence, we consider that it would be sufficient to provide a 2DOF input device (e.g., mouse or P5 glove in a 2D mode) to perform relatively simple selection tasks.

This research is part of a larger project aimed to develop a multimodal visualization environment allowing clinicians to alternate desktop and virtual realities in an adaptive manner while performing medical exploration tasks (Zudilova-Seinstra 2006). Future studies will consider more complex display and device configurations, as well as the importance of stereopsis in noticing selection and positioning challenges.

References

Adame IM, van der Geest RJ, Wasserman BA, Mohamed M, Reiber JHC, Lelieveldt BPF (2004) Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images. MAGMA (Magn Reson Mater Phys Biol Med) 16(5):227–234

Berg C, Hertzog C, Hunt E (1982) Age differences in the speed of mental rotation. Dev Psychol 18(1):95–107

Bornik A, Beichel R, Kruijff E, Reitinger B, Schmalstieg D (2006) A hybrid user interface for manipulation of volumetric medical data. In: Proceedings of the IEEE symposium on 3D user interfaces, pp 29–36

Botella C, Baños RM, Perpiñá C, Villa H, Alcañiz M, Reym A (1998) Virtual reality treatment of claustrophobia: a case report. Behav Res Ther 36(2):239–246

Bowman DA, Johnson DB, Hodges LF (1999) Testbed evaluation of virtual environment interaction techniques. In: Proceedings of the ACM symposium on virtual reality software and technology, London UK, pp 26–33

Bowman DA, Kruijff E, LaViola JJ Jr, Poupyrev I (2005) 3D user interfaces: theory and practice. Pearson Education, Boston

Brooks F P Jr (1999) What's real about virtual reality? IEEE Comput Graph Appl 19(6):16–27

Conner BD, Snibbe SS, Herndon KP, Robbins DC, Zeleznik RC, van Dam A (1992) Three-dimensional widgets. In: Proceedings of the 1992 symposium on interactive 3D graphics. ACM Press, New York, NY, USA, pp 183–188

Cramer HSM, Evers V, Zudilova EV, Sloot PMA (2004) Context analysis to support development of virtual reality applications. Virtual Real 7(3):177–186

Dix A, Finlay J, Abowd G, Beale R (1993) Human computer interaction. Prentice-Hall, New York

Eberts RE (1994) User interface design. Prentice-Hall, Englewood Cliffs

Egan DE (1988) Individual differences in human–computer interaction. In: Helander M (ed) Handbook of human–computer interaction, Elsevier, North Holland, pp 543–568

Egan D, Gomez M (1985) Assaying, isolating, and accommodating individual differences in learning a complex skill. In: Dillon R (ed) Individual differences in cognition, Academic Press, London, pp 173–217

Feng J, Spence I, Pratt J (2007) Playing an action video game reduces gender differences in spatial cognition. Psychol Sci 18(10):850–855

Gabbard JL, Swartz K, Richey K, Hix D (1999) Usability evaluation techniques: a novel method for assessing the usability of an immersive medical visualization VE. In: Proceedings VWSIM'99, pp 165–170

Gagnon D (1985) Videogames and spatial skills: an exploratory study. Educ Commun Technol 33(4):263–275

Hartman NW, Connolly PE, Gilger JW, Bertoline GR, Heisler J (2006) Virtual reality-based spatial skills assessment and its role in computer graphics education. In: Proceedings of the

international conference on computer graphics and interactive techniques (ACM SIGGRAPH 2006), Educators program, article No. 46

He T, Kaufman AE (1993) Virtual input devices for 3D systems. In: Proceedings of IEEE visualization 1993. IEEE Computer Society, San Jose, CA, pp 142–148

Hinckley K, Tullio J, Pausch R, Proffitt D, Kassel DN (1997) Usability analysis of 3D rotation techniques. In: UIST '97: proceedings of the 10th annual ACM symposium on user interface software and technology, New York, NY, pp 1–10

Hodges L, Anderson R, Burdea G, Hoffman H, Rothbaum B (2001) Treating psychological and physical disorders with VR. IEEE Comput Graph Appl 21(6):25–33

Hoffman HG, Patterson DR, Carrougher GJ et al (2001) The effectiveness of virtual reality based pain control with multiple treatments. Clin J Pain 17:229–235

Jacobson J, Redern MS, Furnan JM, Whiney LW, Sparto PJ, Wilson JB, Hodges LF (2001) Balance NAVE: a virtual reality facility for research and rehabilitation of balance disorders. In: Proceedings of virtual reality software technology (VR'01), pp 103–109

Jin W, Lim Y-L, Xu XG, Singh TP, Suvranu DE (2005) Improving the visual realism of virtual surgery. In: Proceedings medicine meets virtual reality, vol 13, pp 227–233

Johnson C (2004) Top scientific visualization research problems. IEEE Comput Graph Appl 4:13–17

Karwowski W (2006) International encyclopedia of ergonomics and human factors. CRC Press, Boca Raton

Krüger A, Irrgang S, Hertel I, Strauss G, Preim P (2007) Comparison of 2D and 3D input devices for virtual endoscopy. In: CURAC 2007, pp 215–218

Larson P, Rizzo AA, Buckwalter JG, van Rooyen A, Kratz K, Neumann U, Kesselman C, Thiebaux M, van der Zang C (1999) Gender issues in the use of virtual environments. Cyberpsychol Behav 2(2):113–124

Leitheiser B, Munro D (1995) An experimental study of the relationship between spatial ability and the learning of a graphical user interface. In: Proceedings of the inaugural America's conference on information systems, Pittsburgh, PA, pp 122–124

Lohman DF (1996) Spatial ability and G. In: Dennis I, Tapsfield P (eds) Human abilities: their nature and assessment. Erlbaum, Hillsdale, pp 97–116

Luursema J-M, Verwey WB, Kommers PAM, Annema J-H (2008) The role of stereopsis in virtual anatomical learning. Interact Comput 20(4):455–460

Martens J-B, van Liere R, Kok A (2007) Widget manipulation revisited: a case study in modeling interactions between experimental conditions. In: Proceedings of the IPT-EGVE symposium, The Eurographics Association, paper 1021

Moise A, Atkins MS, Rohling R (2005) Evaluating different radiology workstation interaction techniques with radiologists and laypersons. J Digit Imaging 18(2):116–130

Myers B, Bhatnagar R, Nichols J, Peck C, Kong D, Miller R, Long AC (2002) Interaction at a distance: measuring the performance of laser pointers and other devices. In: Proceedings of the conference on human factors and computing systems (CHI'2002), Minneapolis, Minnesota, USA, pp 33–40

Passig D, Eden S (2001) Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children. Cyberpsychol Behav 4(6):681–686

Plichta SB, Garzon LS (2008) Statistics for nursing and allied health. Williams & Wilkins, Lippincott

Rizzo A, Buckwalter JG, van der Zaag C (2000) Virtual environment applications in clinical neuropsychology. The handbook of virtual environments. L.A. Erlbaum, New York

Robinson P (2001) Cognition and second language instruction. Cambridge University Press, London

Roessler A, Grantz V (1998) Performance evaluation of input devices in virtual environments. In: Proceedings of the IFIP working group 13.2 conference on designing effective and usable multimedia systems, vol 133, pp 68–73

Salthouse T, Babcock R, Skovronek E, Mitchell D, Palmon R (1990) Age and experience effects in spatial visualization. Dev Psychol 26(1):128–136

Schroeder WJ, Martin K, Lorensen B (2002) The visualization toolkit: an object-oriented approach to 3D graphics, 3rd edn. Prentice-Hall, Englewood Cliffs

Sloot PMA (2000) Simulation and visualization in medical diagnosis: perspectives and computational requirements. In: Marsh A, Grandinetti L, Kauranne T (eds) Advanced infrastructures for future healthcare. IOS Press, Amsterdam, pp 275–282

Stanney KM, Mourant RR, Kennedy RS (1998) Human factors issues in virtual environments: a review of the literature. Presence 7(4):327–351

Strong S, Smith R (2001) Spatial visualization: fundamentals and trends in engineering graphics. J Ind Technol 18(1):1–13

Velez MC, Silver D, Tremaine M (2005) Understanding visualization through spatial ability differences. In: Proceedings of IEEE visualization 2005, Minneapolis, MN, USA, pp 511–518

Vicente K, Hayes B, Williges R (1987) Assaying and isolating individual differences in searching a hierarchical file system. Hum Factors 29(3):349–359

Waller D (2000) Individual differences in spatial learning from computer-simulated environments. J Exp Psychol Appl 8:307–321

Waller D, Knapp D, Hunt E (2001) Spatial representations of virtual mazes: the role of visual fidelity and individual differences. Hum Factors 43(1):147–158

Wingrave C, Tintner R, Walker B, Bowman D, Hodges L (2005) Exploring individual differences in raybased selection: strategies and traits. In: Proceedings of IEEE virtual reality 05, pp 163–170

Zudilova EV, Sloot PMA (2005) Bringing combined interaction to a problem solving environment for vascular reconstruction. Future Gener Comput Syst 21(7):1167–1176

Zudilova-Seinstra EV (2006) Combining desktop and virtual realities: addressing demands of real life clinical environments. Interfaces 67:11–13