Evaluation of quality of experience in interactive 3D visualization: methodology and results

Sylvain Tourancheau[a], Mårten Sjöström[a], Roger Olsson[a], Anders Persson[b], Thomas Ericson[c]

[a] Dept. of Information Technology and Media, Mid Sweden University, Sundsvall, Sweden
[b] Center for Medical Image Science and Visualization, Linköping University Hospital, Sweden
[c] Setred AB, Stockholm, Sweden

ABSTRACT

Human factors are of high importance in 3D visualization, but subjective evaluation of 3D displays is not easy because of a high variability among users. This study aimed to evaluate and compare two different 3D visualization systems (a market stereoscopic display, and a state-of-the-art multi-view display) in terms of task performance and quality of experience (QoE), in the context of interactive visualization. An adapted methodology has been designed in order to focus on 3D differences and to reduce the influence of all other factors. 36 subjects took part in an experiment during which they were asked to solve different tasks in a synthetic 3D scene. After the experiment, they were asked to judge the quality of their experience, according to specific features. Results showed that scene understanding and precision was significantly better on the multi-view display. Concerning the quality of experience, visual comfort was judged significantly better on the multi-view display and visual fatigue was reported by 52% of the subjects on the stereoscopic display. This study has permitted to identify some factors influencing QoE such as prior experience and stereopsis threshold.

Keywords: 3D-visualization, quality of experience, task performance, subjective assessment methodology

1. INTRODUCTION

In the last few years, different kinds of three-dimensional (3D) displays have entered the market. These displays extend the number of recreated depth cues, and so give the viewer a more complete 3D experience. A question is what advantages binocular disparity and motion parallax give when a 3D scene shall be described, analyzed or acted upon. The human visual system (HVS) registers a number of depth cues that give hints about the depth of a scene. These may be classified in monocular or binocular depth cues depending on the number of eyes required to register them. Alternatively, they may be divided into physiological and psychological depth cues depending on whether they are inherent in the physics of the HVS or if they are learnt from visual experiences. All recreated depth cues must be coherent for the viewer to experience a meaningful depth. Failing to do so may cause irritation, headaches or eye strains.

The aim of the present work was to investigate the advantages new 3D displays bring with their additional depth cues. It is well-known that binocular disparity (or stereopsis) is the most efficient depth cue to recreate depth perception. For this reason as well as for technical reasons, most of the 3D displays currently available rely on stereopsis to present depth to observers. It exists many ways to display stereoscopic pictures. Most common systems use special displays and goggles that permit to present independently two different images to the eyes, either simultaneously using polarized glasses or sequentially using shutter glasses. While these systems are very effective to produce immediate depth effect, there is no change in the content as the observer moves around since motion parallax is missing. Though, motion-parallax (or kineopsis) is often said to be the most efficient monocular cue to generate a sensation of relative depth.[1,2] For example, the use of motion-parallax alone with legacy 2D displays can provide a sense of 3D (pseudo 3D)[3] and increase task performance.[4] Although it has been shown that binocular disparity alone provides a more accurate depth discrimination than monocular motion-parallax alone,[5] it is still unclear what are the benefits of combining both stereoscopy and motion-parallax.

Further authors information: (send correspondence to M.S.)
S.T., M.S., and R.O.: [firstname].[lastname]@miun.se (without accent)
A.P.: anders.persson@cmiv.liu.se, T.E.: thomas.ericson@setred.com

This combination is achieved, for instance, by multi-view 3D displays which provide a look-around capability by displaying multiple perspectives of the scene. This is achieving by mounting a parallax barrier or a lenticular lens panel in front of the display to ensure that each pixel will be see only from some specific positions. In majority, these displays are auto-stereoscopic, i.e. users are not required to wear spectacles to see depth. This feature is another advantage of multi-view displays (along with motion-parallax) since it can positively influence ergonomic factors such as quality of experience, visual comfort and visual fatigue.

The experiment presented in this paper aimed to evaluate and compare the influence of two different 3D-displays on task performance and user quality of experience. This evaluation took place in the context of interactive visualization, where users had to understand, analyze and act upon 3D scene. Two different 3D visualization systems were tested. The first one was a time-sequential stereoscopic liquid crystal display (LCD), which adds binocular parallax to the already present monocular depth cues, thanks to the use of active shutter-glasses. It permitted to the viewer to experience real 3D visualization by the mean of stereoscopy. The second one was a state-of-the-art multi-view auto-stereoscopic 3D-display, which supplements the previous depth cues with motion-parallax that implies the possibility to see the scene from different perspectives by moving the position of the observer's head. 36 test persons took part in the experiment, 18 on each visualization system. This paper presents the results of the evaluation of their performance and quality of experience.

The outline of this document is as follows. Section 2 describes the methodology of the experiment, and more precisely the test conditions, the subjects, and the protocol. Section 3 presents the results and discusses the differences between the two visualization systems under test. Finally, Section 4 summarizes the conclusions of this work.

2. METHODOLOGY

2.1 Test conditions

2.1.1 3D visualization systems

Two different visualization systems have been used in this study:

1. The first one consists in a time-sequential stereoscopic liquid crystal display (LCD) together with the NVIDIA 3D Vision^{TM} system.[6] This system presents left and right images sequentially to the observer with a rate of 120 images per second (60 images per second for each eye). This is achieved with liquid crystal shutter-glasses synchronized with the display through infra-red communication.

2. The second visualization system is a time-multiplexed auto-stereoscopic multi-view 3D display using a state-of-the-art scanning slit system.[7,8] Contrary to spatially-multiplexed auto-stereoscopic displays (e.g. using lenticular sheet or parallax barrier in front of the pixels panel) which perform a sampling of the 3D-scene at specific discrete angular positions, this display realizes a uniform sampling of the 3D scene in both spatial and angular dimensions, which permits to present the 3D scene under the form of a *light-field* with a "continuous" motion parallax rather than a set of views. For this experiment, we chose a parameterization with a viewing zone of 36 cm at a viewing distance of 72 cm.

The presentation of 3D scenes was realized thanks to a 3D visualization software dedicated to medical applications. This software performs a real-time rendering of standard CT (computed tomography) medical data sets. The 3D reconstructed volume can then be displayed on one or the other of the two 3D-displays under test.

2.1.2 Displays settings and calibration

The purpose of this evaluation was to assess the influence of 3D features on users' performance and quality of experience. For this reason, display parameters were set as similar as possible on the two visualization systems in order to have homogeneous viewing conditions. For practical reasons, the settings of the stereoscopic LCD were adjusted to match those of the time-multiplexed auto-stereoscopic multi-view display. The resolution of the presentation was set to 1024×768 pixels, which is the native resolution of the auto-stereoscopic display. Since the native resolution of the stereoscopic LCD is 1920×1080 pixels, the presentation was displayed in the center of the display to match the pixels grid and avoid re-sampling. In those conditions, the format of the
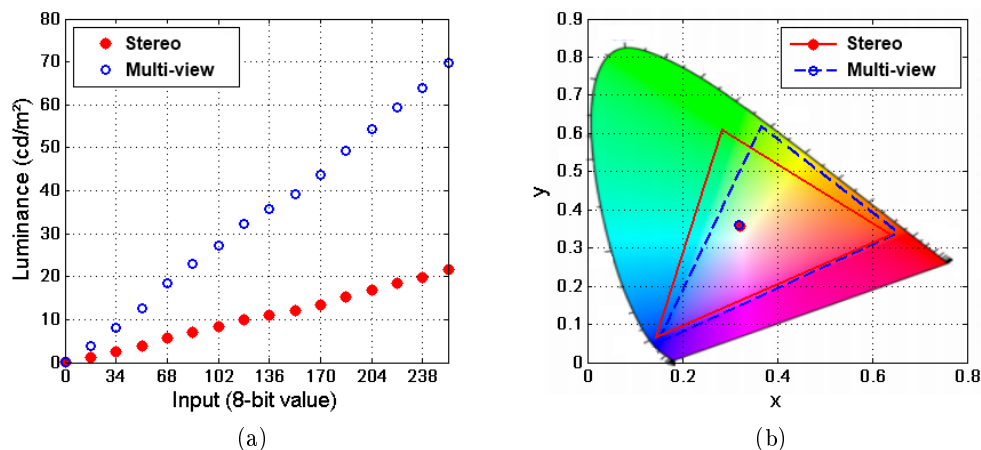
Figure 1: Electro-optical transfer functions (a) and color gamuts and white-points (b) of both displays.

	2D / Stereo	Multi-view
Presentation (pixels)	1024×768	1024×768
format (mm)	272×204	440×330
Viewing distance (mm)	443	717
Black luminance (cd/m²)	0.03	0.05
White luminance (cd/m²)	21.6	69.7
Gamma value	1.0	1.0
White Point (K)	6200	6200

Table 1: Summary of display parameters.

presentation was 440×330 mm on the auto-stereoscopic display, and 272×204 mm on the stereoscopic LCD. The visual resolution of the presentation was set to 30 pixels per visual degree in both configurations. The viewing distance was thus set to 72 cm on the auto-stereoscopic display, and to 44 mm on the stereoscopic LCD. In that conditions, the presentation subtended 34.1°×25.6° of the visual field in both configurations.

Luminance and colors were calibrated in order to be as similar as possible in the two configurations. Again, for practical reasons gamma function and color gamut of the stereoscopic LCD were modified in order to match those of the auto-stereoscopic multi-view display. The electro-optical transfer function was measured on the time-multiplexed auto-stereoscopic 3D display thanks to a Konica-Minolta LS-110 luminance-meter. The gamma value of this display was equal to 1.0 and the white luminance was 69.7 cd/m². The electro-optical transfer function of the stereoscopic LCD display (measured through the shutter-glasses) was then modified accordingly thanks to GretagMacBeth Eye-One Match 3 software used with an X-rite i1 Pro spectrophotometer. However, even with a peak luminance set to its maximum value (355 cd/m²), the luminance of the stereoscopic LCD measured through the active shutter-glasses was only equal to 21.6 cd/m². Figure 1a illustrates these two electro-optical functions. Lighting of the room has been adjusted to compensate this difference (cf. Section 2.1.3). The white-point color temperature of the time-multiplexed auto-stereoscopic 3D display were measured thanks to the X-rite i1 Pro spectrophotometer, and found equal to 6200 K. As previously, GretagMacBeth Eye-One Match 3 software used with an X-rite i1 Pro spectrophotometer was used to calibrate the stereoscopic LCD display accordingly. Figure 1b illustrates the color gamuts and the white points of both systems. It can be observed that color gamuts are different, particularly for the green components. For this reason, only gray stimuli were used in the presentations. Table 1 summarizes displays parameters for each configuration.

	Number of subjects	Average age	Median age	Female ratio (%)
Group 1	18	28.7	26.5	22.2
Group 2	18	27.8	26.5	22.2

Table 2: Average age and female ratio of each group of subjects.

2.1.3 Test room

Tests were carried out in a normalized room, following recommendation ITU-R BT.500-11.[9] The chromaticity of background behind the displays was D_{65}. The room lighting was realized thanks to fluorescent lamps with a color temperature of 6500 K. A particular care was taken in the choice of a high-frequency lighting system to avoid perception of flickering through the shutter-glasses. The lighting intensity was set in order to obtain a background luminance behind the displays equal to 15% of the display peak luminance as recommended. Thus, two different room settings were used to take into account the difference in peak luminance between both configurations: the background luminance was equal to 10 cd/m² when the auto-stereoscopic display was used, and equal to 3 cd/m² with the stereoscopic LCD.

2.2 Test Subjects

2.2.1 Groups of subjects

Thirty-six naïve test subjects took part in this experiment. All of them were employees or students of Mid Sweden University, Sundsvall Campus.

Test subjects were separated in two groups of 18 persons. The first group (Group 1) performed the experiment with the stereoscopic display and the second group (Group 2) with the multi-view display. After this first session, 10 test subjects from Group 1 and 9 subjects from Group 2 performed the test a second time on the other visualization system. This second session took place 12 days in average after the first session.

A particular care was taken in order to have group of subjects as homogeneous as possible regarding the age distribution and the gender proportion. These statistics are given in Table 2.

2.2.2 Vision check-up

Subjects' vision was tested preliminary to the first test session. Visual acuity of both eyes (with correction when necessary) was measured and only subjects with at least 0.9 (decimal notation) for each eye were allowed to participate in the experiment. Subjects with color blindness were screened thanks to Ishihara plates. Moreover, the binocular parallax (stereopsis) threshold, or stereoacuity, of subjects was measured using to the TNO stereopsis test. It corresponds to the minimum binocular disparity they can perceive as depth.

2.3 Test protocol

The goal of this experiment was to evaluate the influence of the visualization system on task performance and user quality of experience. As stated previously in Section 2.1.1, the two visualization systems were used along with a 3D-visualization software dedicated to medical applications. This software is used for medical analysis and diagnosis and provides a lot of tools to interact with and modify the 3D volumes. In this experiment, subjects' interactions were limited to rotations and only the marker tool, which permits to place a marker at a specific location of the scene, was made available. Prior to the experiment, subjects were asked to solve a training test, which was designed in order to teach team how to use the software (in the particular case of this experiment).

2.3.1 Test object

It was decided to use a real artificial object as an imaging phantom, the geometry of which being known, so that the ground truth was perfectly defined. Our choice was directed towards a puzzle-game which consists in a plastic polyhedron traversed by a dozen of elastic strings the extremities of which can be moved in order to create specific patterns (Figure 2a). This phantom was scanned by Computed Tomography resulting in 475 slices with a 512×512 pixels resolution. The 3D reconstruction (Figure 2b) of the object was performed by the visualization software at a resolution adapted to the visualization system under test (cf. Table 1).
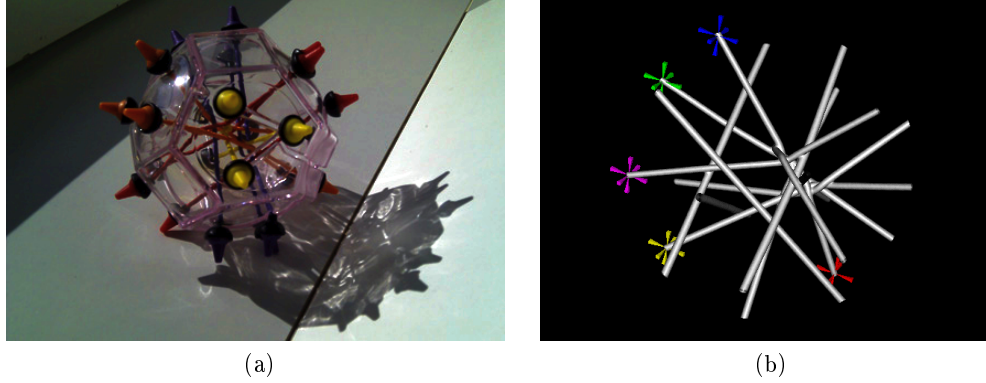
(a)   (b)

Figure 2: Puzzle-game used as a phantom (a), 3D reconstruction of the inside from CT (b).
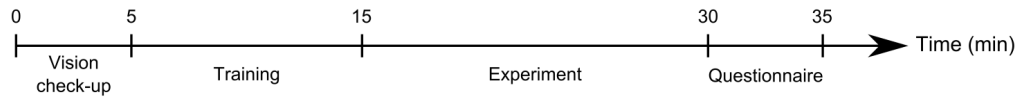

Figure 3: Time-line of the experiment (first session).

Three different configurations of the puzzle-game were scanned. In the first one, the twelve elastic strings were positioned at the initial location, with no entanglements between them. In the second and the third configurations, strings were moved and tangled in order to create some "knots" inside the polyhedron. Finally, the plastic polyhedron surface of the puzzle-game was digitally removed in order to focus the tasks on the elastic strings' position. The rendering presets of the visualization software were set in order to avoid any chrominance differences between systems (because of the differences in color gamut which had been measured previously, cf. section 2.1.2). During the experiment, some color markers were added to the scene in order to facilitate instructions to the subjects (by identifying specific strings for example, cf. Figure 2b).

2.3.2 Tasks

The experiment consisted in the resolution of 13 tasks involving the 3D-visualization of the test object. Table 3 gives the instruction for each task. The tasks were presented to all test subjects with the same chronological order from T1 to T13. Two different types of tasks can be distinguished: static tasks and dynamic tasks according to whether subjects are allowed or not to move the volume. For each type, tasks can be classified in three categories regarding the actions which were involved: scene understanding, precision (marker placing), and depth perception. The latter could be performed only for static tasks, since moving the scene would change depth order. Table 6 gives the classification of each task.

2.3.3 Procedure

Time-line

Figure 3 gives the time-line of the first test session from the moment subjects enter the test-room to the moment they leave. The average times for each part of the test are also given. Subjects vision was first checked, then subjects were introduced to the visualization system they would use, and the training session was performed (average length of the training session was 10 minutes). The experiment itself was then done within an average time of 15 minutes and subjects were finally asked to fill a questionnaire about quality of experience, comfort and fatigue at the end of the session.

The second session consisted of the same stages except that the vision check-up was not performed again. The average period of time between the first and the second session was 12 days, with a minimum 6 days, and a maximum of 20. It was decided to ask the subjects to perform the training again during the second session so that they had the same memory of the software.

Task #	Instruction	Interaction
T1	How many strings are there in the scene?	No
T2	Which of the three marked strings is the nearest to you?	No
T3	Which of the three marked strings is the farthest to you?	No
T4	How many strings are there in the scene?	Yes
T5	Find the place where the blue string and the yellow string are the closest to each other. Place a marker at half the distance between strings at this place.	Yes
T6	Find the place where the blue string and the yellow string are the closest to each other. Place a marker at half the distance between strings at this place.	No
T7	How many strings are behind the plane defined by the four markers?	No
T8	How many strings are in front of the plane defined by the four markers?	No
T9	How many strings have no point of contact with others?	Yes
T10	Which marker is at the other end of the red-marked string?	No
T11	Which marker is at the other end of the red-marked string?	No
T12	Which marker is at the other end of the red-marked string?	Yes
T13	Find the place where the blue string and the red string are the closest to each other. Place a marker at half the distance between strings at this place.	Yes

Table 3: Description of the 10 tasks.

	Static	Dynamic
Marker placing	6	5, 13
Scene understanding	1, 10, 11	4, 9, 12
Depth perception	2, 3, 7, 8	

Table 4: Classification of the tasks.

Training

The training session was designed with a growing difficulty in order to teach subject how to use the visualization software, particularly how to rotate the ball in the three dimensions using the click-and-drag mouse feature, and how to place a marker at a specific location of the scene. Since most of the subjects did not use English as a mother tongue, the training was also the occasion for the operator to precise the terms used in the instructions of the experiment such as depth-related terms (in front of, behind, nearest, farthest, etc.) and geometric terms (distance, plane, etc.).

Stimuli used in the training session consisted of markers and simple geometric objects such as squares and circles. The phantom was not used in the training session in order to ensure that every subject starts the experiment with no prior knowledge of it.

Recordings

Solving time was recorded independently for each task. Subjects were asked to proceed as follows: read the instruction first, start the timer, solve the task, stop the timer, and finally write down their answer when necessary.

For most of the tasks, subjects were asked to write down their answers on a paper sheet. Result then takes the form of a binary number (correct or not). For tasks 5, 6, and 13, the position of the marker set by the subjects was retrieved and the Euclidean distance to the correct position was computed. In that case, result takes the form of a visual angle expressed in minutes of arc as measured at the display plane.

Score	Adjective
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 5: 5-grade subjective quality scale used to assess the quality of experience at the end of the test.

Questionnaire

At the end of the experiment, subjects were asked to fill a questionnaire in order to evaluate their quality of experience. Four questions have been asked within this questionnaire:

- Q1: How would you rate the global quality of the visualization system?

- Q2: How would you rate the 3D experience delivered by the display? ...

- Q3: How would you rate the visual quality of the display?

- Q4: How would you rate your visual comfort during the experience?

- Q5: Do you experience visual fatigue (eye strain, headache, etc.) at the end of the experiment? (If yes, please comments)

For the four first questions, subjects gave a score according to a 5-grade quality scale often used in subjective quality evaluation (Figure 5a). For the last question, a Yes or No answer was expected and subjects were asked to describe their sensations if they did experience visual fatigue.

3. RESULTS

3.1 Task performance

3.1.1 Solving times

Figure 4 presents a comparison of tasks' solving times between the multi-view display and the stereoscopic display, for both test sessions. Solving times are presented as the mean values with 95% confidence intervals. Figure 4a gives results of the first session and concerns all test subjects (18 test subjects on each display) and Figure 4b gives results of the second session and concerns only test subjects who did the test a second time (10 and 9 test subjects on each display). For both sessions, tasks' solving times on the multi-view display and tasks' solving times on the stereoscopic display are very similar. There are no significant differences at the 5% level between displays for any of the tasks.

When compared between sessions, solving times were lower during the second session, as expected. If these solving times are compared only for the 19 test subjects who did the experiment twice, it can be observed that those who solved the tasks on the multi-view display first were 22% faster in average on the stereoscopic display, and those who solved the tasks on the stereoscopic display first were 20% faster in average on the multi-view display.

3.1.2 Tasks responses

Results of the tasks are presented in Figure 5 for each category of tasks (cf. Table 4). Results are compared between both visualization systems, for all subjects and for both test sessions in Figure 5a. Figure 5b and Figure 5c presents the results respectively of the first session and of the second session, only for the subset of 19 subjects who performed the experiment twice. Results of the tasks implying scene understanding (tasks 1, 4, 9, 10, 11, and 12) and implying depth perception (tasks 2, 3, 7, and 8) are presented as the percentage of correct responses.

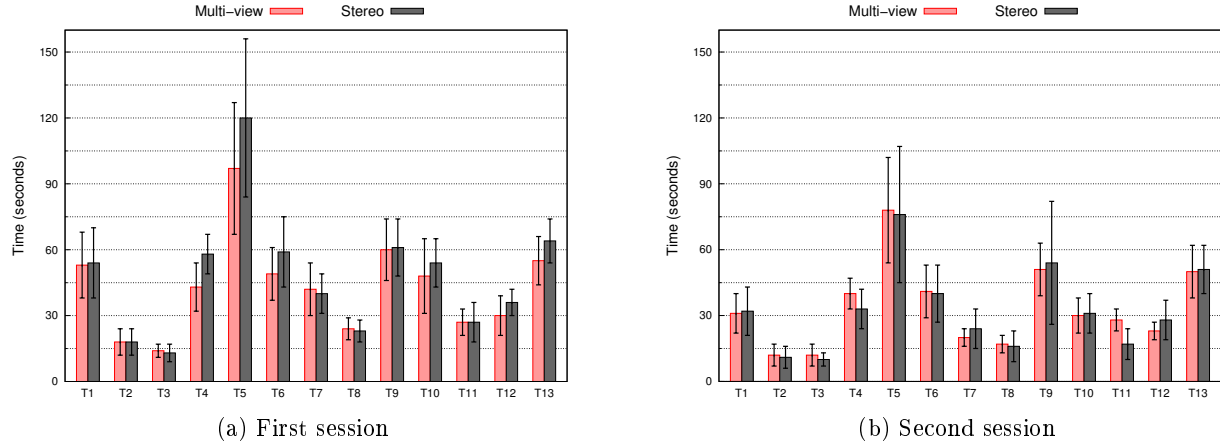(a) First session   (b) Second session

Figure 4: Comparison of tasks' solving times between the multi-view display and the stereoscopic display, for the first session (a) and the second session (b). Error bars depict 95% confidence intervals.

Results of the precision tasks (tasks 5, 6, and 13) are presented as the mean distance between the marker placed by the test subjects and the correct position, in minutes of arc.

Averaged results from both test sessions and for all test subjects (Figure 5a) show no significant differences (at the 5% level) between both displays. Results are very similar concerning depth perception tasks, with an average of about 85% of correct responses for both displays. However, difference is found close to statistical significance for tasks dealing with scene understanding: in average 65% of responses were correct on the multi-view display and only 50% on the stereoscopic display. In the same way, the accuracy of subjects is better on the multi-view display (average distance of about 30 arcmin) than on the stereoscopic display (average distance of about 1 degree) with a difference close to statistical significance.

If only the subset of 19 subjects who performed the experiment twice is considered, differences in task performance between both sessions can be observed (Figure 5b-c). For both categories of tasks mentioned before (scene understanding and precision), task performance during the second session was significantly better on the multi-view display. When compared between sessions, results show that subjects who did the experiment on the stereoscopic display significantly improved their performance during the second session on the multi-view display. On the contrary, subjects who did the experiment on the multi-view first obtained worse results during the second session on the stereoscopic display. Again, no specific differences between displays and sessions are observed for the four tasks dealing with depth perception.

3.1.3 Discussion

From a global point of view, users' performance was better on the multi-view display than on the stereoscopic display, for tasks asking for precision and scene understanding. This is particularly true for the second test session: subjects significantly improved their performance when passing from the stereoscopic display to the multi-view display, but the transition from the multi-view display to the stereoscopic display slightly decreased their performance. For tasks involving depth perception only (which were all static tasks), no significant differences in the performance were observed between both displays. This is a surprising results since the multi-view display features a look-around capability which is particularly efficient (with respect to the stereoscopic display) for static situations. This might be due to a difficulty level which was too low for these four tasks and so not distinguishing enough.
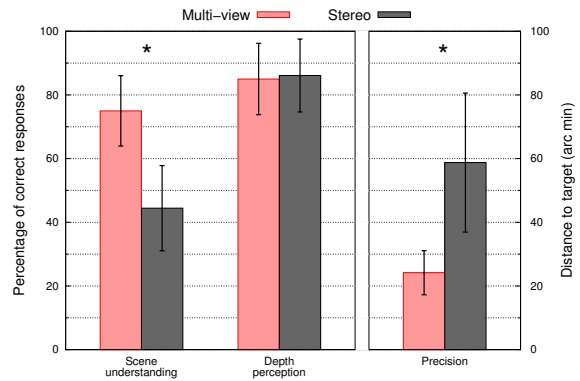
Concerning tasks' solving times, no significant differences were observed between both displays. In other terms, test subjects performed better on the multi-view display while spending in average the same amount of time on the tasks.
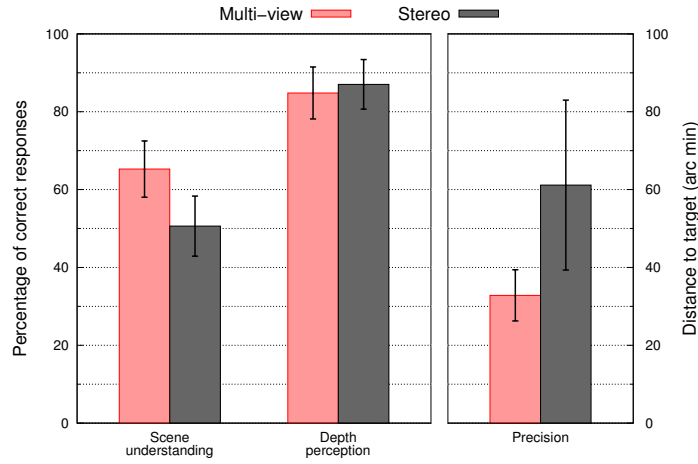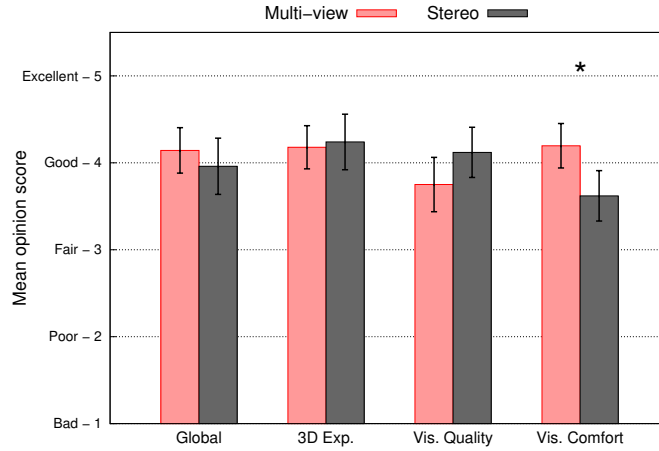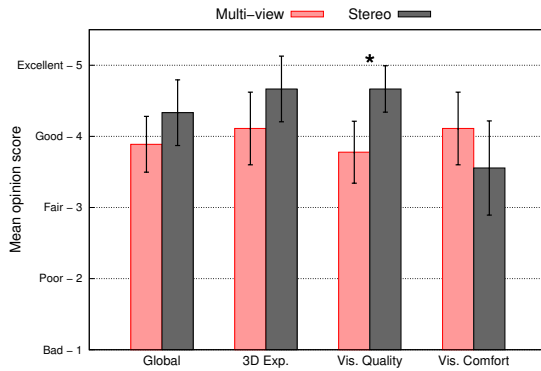
(a) Both sessions.
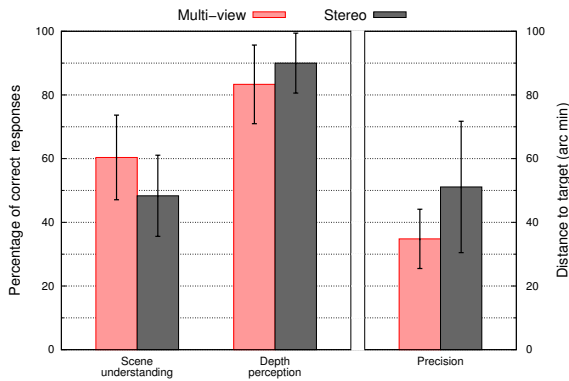
(b) First session.

(c) Second session.

Figure 5: Comparison of task performance on both displays for the three categories of tasks described in Section 2.3.2. (a) Results are averaged over all test subjects and both sessions. (b) Results are presented only for the subset of subjects who performed twice, for the first session. (b) Results are presented only for the subset of subjects who performed twice, for the second session (c). Error bars depict 95% confidence intervals. Asterisks indicates significant difference at the 5% level.

(a) Both sessions.

(b) First session.
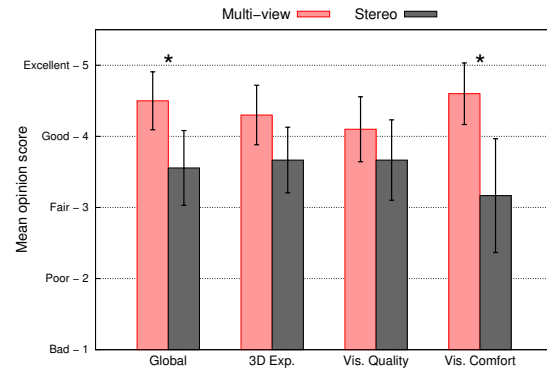
(c) Second session.

Figure 6: Comparison of quality of experience evaluation on both displays. (a) Results are averaged over all test subjects and both sessions. (b) Results are presented only for the subgroup of subjects who performed twice, for the first session. (b) Results are presented only for the subgroup of subjects who performed twice, for the second session (c). Error bars depict 95% confidence intervals. Asterisks indicates significant difference at the 5% level.

3.2 User experience

3.2.1 Quality of experience

The results of the subjective evaluation of the quality of experience are presented and compared between both visualization systems in Figure 6. Figure 6a presents the mean opinion scores (MOS) for all test subjects and averaged over both test sessions, for the four features described in Section 2.3.3 (questions Q1 to Q4). Results have been found equivalent (no statistical significance at the 5% level) between both displays for the three first features (global quality, 3D experience, and visual quality), with an average score around "Good" on each display. Visual comfort was found significantly better (p=0.028) on the multi-view display rather than on the stereoscopic display, with a difference of about 0.5 points.

Figure 6b-c presents the same results only for the subgroup of 19 test subjects who performed the experiment twice. Results are presented independently for each session. Significant differences are observed between both sessions. After the first session the visual quality has been judged significantly better on the stereoscopic display, while the three other features were found equivalent. After the second session, the same subjects – that time
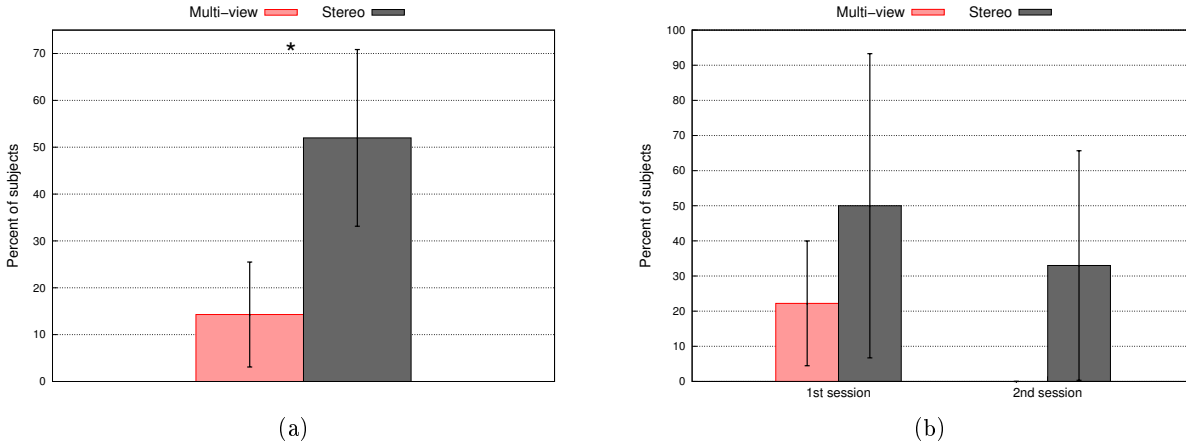
Figure 7: Percentage of test subjects who experienced visual fatigue at the end of the experiment, for both sessions and all test subjects (a), and separately for first and second session, only for the subgroup of 20 subjects who performed the experiment twice. Error bars are 95% confidence intervals. Asterisks indicates significant difference at the 5% level.

performing on the other display – judged the visual comfort and the global quality significantly better on the multi-view display (with a difference of 1.5 points and 1 point respectively), while visual quality and 3D experience were equivalent.

3.2.2 Visual fatigue

Figure 7a presents the percentage of test subjects who experienced visual fatigue at the end of the experiment, for all test subjects and both sessions. A significant difference in terms of visual fatigue is observed (p=0.016) in favor of the multi-view system: 51% of the subjects experienced visual fatigue at the end of the experiment on the stereoscopic display against only 13% on the multi-view display. Figure 7b presents the same results but this time compared between both test sessions, and only for the subgroup of 19 subjects who did the experiment twice. The same tendency is observed (visual fatigue reported on the stereoscopic display in majority) but with no significant differences due to the larger extent of confidence intervals.

Subjects who reported visual fatigue were asked to comment their sensations and to express what they felt. The feedback of the test subjects was categorized in three groups as following:

1. Eye strain, tiredness: reported as a feeling of strain in the eyes, potentially led to a blurry vision

2. Difficulties to fuse views (on the stereoscopic display only): reported as the apparition of a double image and the necessity to concentrate hard to fuse views and experience 3D.
   (It is hard to know if this is caused by a temporary diplopia due to tired eye muscles, or if this is, more likely, the consequence of the important stereoscopic crosstalk that can be perceived on the display under test; it might also be a combination of both.)

3. Dizziness, vertigo, etc: few subjects reported such physical effects, on the stereoscopic display only

Figure 8 illustrates the percentage of comments which can be classified in each group.

3.2.3 Discussion

From a global point of view, both 3D-visualization systems obtained equivalent mean opinion scores in terms of global quality, 3D experience, and picture quality. Nonetheless, visual comfort was found significantly better on the multi-view display (1 point better on a 5-grade scale). This was confirmed by the fact that half of the observers experienced visual fatigue on the stereoscopic display, while only 13% of them did on the multi-view system.
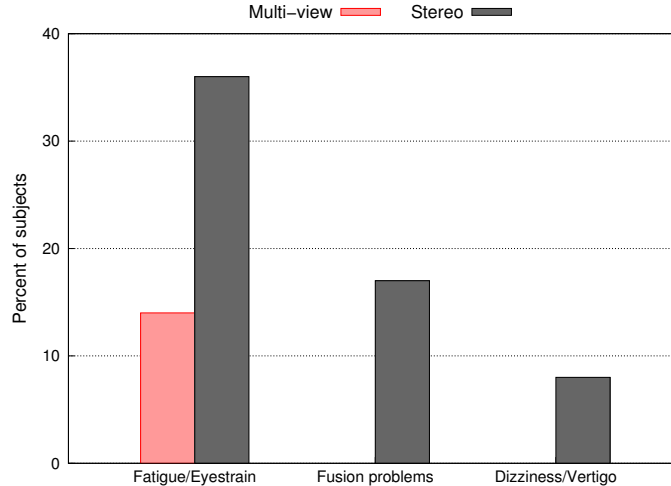
Figure 8: Problems reported by subjects for each display, categorized in three types of discomfort: visual fatigue and eyestrain, difficulties to fuse views. Results concerned both sessions and all test subjects.

When compared between both sessions, the results of this evaluation were found very different suggesting that the experience acquired by the test subjects during the first test session influenced their judgement during the second test session. Thus, the 10 subjects who did the first test session on the stereoscopic display gave an average score of around 4.5 to visual comfort on the multi-view display during the second session, while inversely the 9 test subjects who did the first session on the multi-view display judged the visual comfort as only "Fair" (score around 3) on the stereoscopic display during the second session. These observations suggest that subjects' prior experience is an important factor influencing the evaluation of the quality of experience in 3D-visualization. Users' comfort highly depends on what they expect to be a comfortable experience and this expectation will vary according to previous experiences. This is confirmed by results concerning visual fatigue: after the first session 32% of the test subjects declared that they suffered from visual fatigue, while at the end of the second session, only 16% of the same 20 subjects did so.

Characteristics of binocular vision are another factor which can affect the visual comfort experienced by test subjects. The influence of subjects' stereoacuity (binocular parallax threshold, see Section 2.2.2) on the visual comfort subjective score was investigated. Figure 9 presents visual comfort mean opinion scores on both 3D-displays for subjects with a stereoacuity of 60 arcsec or less and for subjects with a stereoacuity of 120 arcsec or more. It can be observed that the visual comfort on the stereoscopic display for test subjects with a stereoacuity of 120 arcsec or more was significantly lower than for test subjects with a lower threshold (or higher stereoacuity). No significant difference was found on the multi-view display. This result seems to indicate that a lower stereoacuity can increase discomfort on displays which rely only on binocular parallax.

These results point out the necessity of reviewing subjects' vision, background, and prior experience, in order to be sure to gather homogeneous groups of subjects. Particularly, if one wants to compare different 3D-visualization systems, it is important to build groups of test subjects with a similar distribution of stereopsis threshold. Furthermore, using the same groups of subjects to assess different 3D-systems sequentially (as it was done in this study) might be hazardous since test subjects will learn from the first experience and change their expectations accordingly. To avoid this issue, one could exclusively use some expert subjects, who are familiar with the 3D-displays technologies under test. Another solution can be to compare 3D-visualization systems simultaneously, but this might be difficult or impossible for technical reasons.

4. CONCLUSION

This study aimed to compare task performance and quality of experience on two different 3D-visualization systems: a state-of-the-art auto-stereoscopic multi-view display and a time-sequential stereoscopic LCD. 36
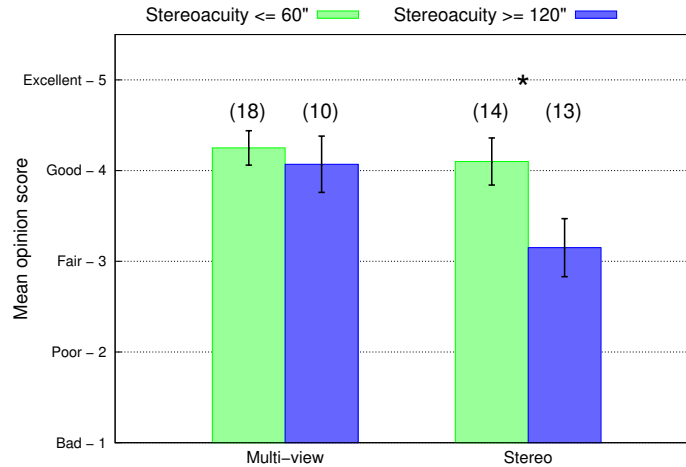
Figure 9: Influence of stereopsis threshold on the score of visual comfort reported by test subjects. MOS are presented for both displays separately for subjects with a stereopsis threshold of 60 arcsec or less, and for subjects with a stereopsis threshold of 120 arcsec or more. Error bars are 95% confidence intervals. Asterisks indicates significant difference at the 5% level. Numbers between parentheses indicates the number of test subjects for each condition.

naïve test subjects took part in a subjective experiment in which they were asked to solve some tasks in a 3D scene, on both displays. Results demonstrate that users' performance was significantly better on the multi-view display for tasks dealing with scene understanding and asking for precision. For tasks involving depth perception only (which were all static tasks), no significant differences in performance have been observed between both displays. Tasks' solving times were found equivalent on both 3D-displays. In other terms, test subjects performed better on the multi-view display while spending the same amount of time on the tasks.

At the end of the experiment, test subjects were asked to assess the quality of their experience during the test. This evaluation showed that both 3D visualization systems obtained equivalent mean opinion scores in terms of global quality, 3D experience, and picture quality. However, visual comfort has been found significantly better on the multi-view display (1 point better on a 5-grade scale), which is confirmed by the fact that half of the observers experienced visual fatigue or eye-strain on the stereoscopic display, while only 13% of them did on the multi-view system. However, it is observed that the results of this evaluation were very different between the first and the second session, suggesting that the experience acquired by the test subjects during the first test session influenced their judgement during the second session. This is confirmed by results concerning visual fatigue: after the first session 36% of the test subjects declared that they suffered from visual fatigue, while at the end of the second session, there were only 16%.

These observations show that subjects' prior experience is particularly important in the quality of experience in 3D-visualization. Users' comfort highly depends on what they expect to be a comfortable experience. Results also suggest that subjects binocular vision characteristics (stereopsis threshold in that case) might have an influence on the visual comfort. These results point out the necessity of reviewing subjects' vision, background, prior experience, in order to be sure to build homogeneous groups of subjects.

5. ACKNOWLEDGMENTS

REFERENCES

[1] J. J. Gibson, *The perception of the visual world*, Houghton Mifflin, New York, 1950.

[2] C. Caudek and D. R. Proffitt, "Depth perception in motion parallax and stereokinesis," *Journal of Experimental Psychology* **19**(1), pp. 32–47, 1993.

[3] C. Zhang, Z. Yin, and D. Florêncio, "Improving depth perception with motion parallax and its application in teleconferencing," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing, MMSP '09*, pp. 1–6, 2009.

[4] C. Ware and G. Franck, "Evaluating stereo and motion cues for visualizing information nets in three dimensions," *ACM Transactions on Graphics* **15**(2), 1996.

[5] J. Faubert, *Three-Dimensional Television, Video, and Display Technologies*, ch. 16. The Influence of Optical Distortions and Transverse Chromatic Aberration on Motion Parallax and Stereopsis in Natural and Artificial Environments, pp. 359–396. Springer, 2002.

[6] NVIDIA, "Nvidia 3D Vision^{TM} system - Overview." http://www.nvidia.com/object/3d-vision-main.html.

[7] O. Cossairt, C. Møller, A. L. Travis, and S. A. Benton, "Novel view sequential display based on DMD^{TM} technology," in *Stereoscopic Displays and Virtual Reality Systems XI, Proceedings of SPIE-IS&T Electronic Imaging* **5291**, 2004.

[8] C. Møller, *Scanning Slit 3D Displays*. PhD thesis, University of Cambridge, 2005.

[9] ITU, "Methodology for the subjective assessment of the quality of television pictures," recommendation ITU-R BT.500-11, International Telecommunication Union, 2004.