DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

INVESTIGATING INTERMITTENT STEREOSCOPY: ITS EFFECTS ON PERCEPTION AND VISUAL FATIGUE

Ari A. Bouaniche, Laure Leroy; Laboratoire Paragraphe, Université Paris 8; Saint-Denis; France

Abstract

In a context in which virtual reality making use of S3D is ubiquitous in certain industries, as well as the substantial amount of literature about the visual fatigue S3D causes, we wondered whether the presentation of intermittent S3D stimuli would lead to improved depth perception (over monoscopic) while reducing subjects' visual asthenopia. In a between-subjects design, 60 individuals under 40 years old were tested in four different conditions, with head-tracking enabled: two intermittent S3D conditions (Stereo @ beginning, Stereo @ end) and two control conditions (Mono, Stereo). Several optometric variables were measured pre- and post-experiment, and a subjective questionnaire assessing discomfort was administered.

Our results suggest a difference between simple scenes (containing few static objects, or slow, linear movement along one axis only), and more complex environments with more diverse movement. In the former case, Stereo @ beginning leads to depth perception which is as accurate as Stereo, and any condition involving S3D leads to more precision than Mono. We posit that the brain might build an initial depth map of the environment, which it keeps using after the suppression of disparity cues. In the case of more complex scenes, Stereo @ end leads to more accurate decisions: the brain might possibly need additional depth cues to reach an accurate decision. Stereo and Stereo @ beginning also significantly decrease response times, suggesting that the presence of disparity cues at task onset boosts the brain's confidence in its initial evaluation of the environment's depth map. Our results concerning fatigue, while not definitive, hint at it being proportional to the amount of exposure to S3D stimuli.

Introduction

Virtual reality has nowadays become a staple for certain industries. Because virtual environments are shown onto 2D screens, the immersion must compensate for the absence of depth if perception is to be as accurate as possible.

It is well-known that, to solve the underconstrained inverse problem of inferring a third dimension out of two flat retinal images, information derived from several cues intervenes. Some cues are static and monocular while others involve proprioception (accommodation and vergence), subject movement (motion parallax) or binocular vision (stereopsis). Another possible classification for depth cues exists as well, according to whether they are pictorial in nature. Non-pictorial cues (which originate from ocular or physiological information, and are thus often referred to as "primary cues") include the aforementioned cues of accommodation and vergence, motion parallax, and binocular disparity.

This latter cue is today vastly used and studied in the domain of virtual reality. Still, two things are quite surprising: a review of the research dealing with the perceptions of depth, shape and distance in virtual environments does not clearly indicate that the use of binocular disparity yields a better performance, and a cursory look at the literature concerning the viewing of stereoscopic 3D (henceforth abbreviated S3D) content leaves little doubt as to some negative consequences on the visual system, at least for a certain part of the population.

While we do not question the utility of S3D and its use in immersive environments in this paper, we aim to position ourselves from the point of view of user safety: if, for example, the use of S3D is a given in a company setting, can an implementation of intermittent horizontal disparities aid in depth perception (and therefore in task completion) while impacting the visual system less negatively than classic S3D contents?

Related work

Binocular disparity and the perception of depth

The cues mentioned in the introduction do not all and always have the same importance in depth decision-making. Cutting and Vishton [1] have produced a graph plotting the just-discriminable depth thresholds as a function of the log of distance from the observer: binocular disparity is most effective in the personal space (0-2 m), but also affects the action space (2-30 m). They however deem their graphic representation "idealized" insofar as it could be dramatically altered by subject stereoweakness (or worse, stereoblindness, a condition affecting 6-8% of the population [2]), therefore often meaning that the effectiveness of binocular disparity is restricted to little more than the personal space, when considering the statistic that more than 20% of the population has some form of binocular anomaly impacting their stereoacuity [3].

In such a context, the literature is rather conflicted in substantiating the use of S3D. Some research does cite stereopsis as a valuable cue for depth perception [4] [5] [6] [7] [8]; there are however several studies which fail to ascertain that disparity is, by itself, useful in depth, shape or distance perception: in most cases, the implementation of head tracking (allowing subjects the use of motion parallax, in conjunction with S3D or not) is a much more helpful cue [9] [10] [11]. Indeed, to be found significant in aiding task completion, stereopsis has to be 'isolated': while testing for its effectiveness, experiments make sure that other depth cues are suppressed. This therefore seems to suggest that disparity is a rather weak cue, potentially aiding depth perception, but rather easily overridden by other, more powerful cues such as motion parallax. The synergetic nature of stereopsis and motion parallax is nonetheless worthy of notice: several studies report that the association of both cues is more effective than the sum of both their respective powers when taken separately [6] [12].

Even though there seems to be a lack of clear evidence as to its helpfulness in appreciating depth, S3D viewing has been largely adopted, as stated before. Its tendency to stress users' visual systems and occasion visual fatigue and discomfort has, as well, been the subject of vast amounts of research.

Visual fatigue and discomfort arising from S3D viewing

Very often, visual fatigue (or asthenopia) and visual discomfort have been considered one and the same. Lambooij et al. however make a difference: fatigue should be measured objectively while discomfort is subjective [13]. Still, they stress that subjective discomfort should be considered as part and parcel of overall visual fatigue: a change in objective measurements only does not necessarily imply fatigue, as it could indicate a mere adaptation.

However much visual fatigue and discomfort have been investigated, there is as of yet no clear definition of the two terms, although all descriptions more or less concur to suggest that they are a combination of several symptoms and their underlying causes. Lambooij et al. propose that visual fatigue is the "physiological strain or stress resulting from exertion of the visual system", which would suggest that it is proportional to an increase of the load imposed on the oculomotor system [13]. The viewing of S3D contents does indeed create several situations which entail excessive demand on the visual system:

Accommodation and vergence mismatch

Accommodation-vergence mismatch has been extensively investigated in prior research. Indeed, it consists in a decoupling of the two cross-linked capabilities involved when fixating on an object: instead of both accommodating and verging toward the same point, the eyes accommodate on the screen but verge to the point in space where the simulated object is located. This puts an unnatural strain on the visual system: indeed, fixating on a depicted S3D object requires the same amount of accommodation (onto the fixed display) regardless of the amount of vergence (onto the perceived object). The changed demand line in the case of any S3D object fixation illustrates this strong constraint: a large part of the resulting demand on the oculomotor system when viewing S3D contents is located outside of Percival's and/or Sheard's zones of comfort [14].

Large horizontal disparities

Large horizontal disparities interact negatively with the accommodation-vergence mismatch: this is a direct consequence of the decoupling of vergence and accommodation. In this context, should the depth at which an object is perceived (through its disparity) exceed the depth of field (the range of distances for which an object is perceived in sharp focus, given the fixed accommodation on the display), Lambooij et al. report that the accommodative response is suppressed regardless of vergence, giving way to three possible error scenarios: either accommodation remains and fusion is lost, resulting in diplopia, or accommodation is lost and the image is therefore blurred, or a combination of both [13]. Although Wöpking observed fusion for disparities of 140 arc min (provided that the image be heavily blurred), he states that the literature traditionally recommends not exceeding a 70 arc min disparity [15].

Sharp focus

An unnatural focus on whole scenes often occasions extremely high spatial frequencies, and elicits fusion efforts even though the disparities are too large, thereby contributing to visual strain, all the more so as spatial frequencies affect the size of Panum's fusional area directly [13]. In natural viewing conditions, however, objects which are not fixated appear blurred to some extent, which lowers their spatial frequency contents, making it easier for the visual system to fuse those objects without experiencing much strain [16] (due to the increase in size of Panum's area with lower frequencies).

Disparity changes

Frequent disparity changes (from crossed to uncrossed and vice-versa) seem to elicit an excessive demand on the accommodation-vergence linkage [13], and potentially have an even greater negative influence on visual comfort than the amount of disparity itself [17].

Other factors

Other factors are also mentioned in the literature, amongst which vertical disparity incoherence (such disparities are very often ignored in S3D contents, mostly generated with parallel cameras, to the sole profit of horizontal disparity cues [18]) and visuo-vestibular incoherence (when the contents suggests motion while subjects are still, potentially leading to motion sickness [19]).

Beyond their definition, there is as of yet no single research method allowing the evaluation of visual fatigue and discomfort [20]. In numerous works, however, the assessment of fatigue and discomfort has been twofold, and follows the distinction between an objectively measurable visual fatigue and a subjectively appreciated discomfort [13] [20].

Numerous studies have shown significant alterations in several optometric variables. The altered oculomotor parameters heralding asthenopia which are most commonly observed in the literature are: stereoacuity, and accommodative and vergence responses [16] [21] [22]. Others, however, find no significant alteration to an array of optometric variables [23] or non-significant correlations of variable alteration with visual discomfort [24]. It is possible that the somewhat conflicting results of these studies are due to some people's being more prone than others to visual fatigue, and therefore to an alteration of optometric variables after S3D viewing [3] [19] [24]. In spite of a scarcity of data on the matter, it seems that non-strabismic binocular disorders do not prevent stereopsis but predispose subjects to visual fatigue in S3D context; such disorders could affect up to 20% of the population3. Moreover, accommodative dysfunctions (which could affect more than 15% of the general population) are much more frequent (about 60%) in subjects with binocular disorders [25]. One can easily deduce that such subjects who combine both binocular and accommodative dysfunctions will be likely to experience visual fatigue when viewing S3D contents. It is therefore essential that some method of testing be devised to screen subjects susceptible to visual fatigue due to S3D without having to implement complex testing of optometric variables, which implies equipment not readily available in all research facilities. Read and Bohr have not managed to find any significant correlations between any testable optometric / orthoptic variable and proneness to asthenopia in their rather large pool of subjects [19]. Lambooij et al., on the other hand, recommend using the Wilkins Rate of Reading test and compare the ratio obtained between a 2D and an S3D context: it seems that the ratio is a good predictor of subjects' binocular statuses [25].

As to the assessment of visual discomfort through subjective questionnaires, it is also very diverse in the literature, either when

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

considering the symptoms evaluated or the format. Criteria evaluated (eye burning, eye ache, eye strain, eye irritation, tearing, blur, double vision, headache, dizziness, and nausea) are rather consistent throughout studies [3] [23] [25] [26] [27], as are respondents' answers, as pointed out by Read and Bohr reviewing previous research: reported symptoms following S3D viewing are mostly related to headache and eyestrain [19]. Criteria classification is however much more varied.

As far as evaluation methodology goes, most studies use Likert-type scales (mostly 5-, sometimes 7-point) in their subjective assessment questionnaires: such scales seem to have become standard assessment procedure [3] [14] [19] [23] [25]. Yet some have chosen to use continuous rating scales [20], which are confirmed to be preferable because they produce less noisy data sets [28].

Our brief review of the literature highlighted that binocular disparity was a rather weak cue and that S3D viewing was tiring for the oculomotor system of at least the most susceptible subjects, occasioning fatigue and discomfort. In the case of work-related S3D use, when S3D is not optional nor is used for recreation, it is important to safeguard the visual system of the workers involved if they are to keep doing their job. This led us to consider an intermittent use of stereoscopy and devise the following experiment.

Experiment

Subjects

The experiment was taken by 60 willing subjects, from 15 to 38 years of age ($M = 27.75$, $SD = 4.949$). The experiment followed an inter-subject design, as it was considered extremely impractical to ask the same subject to come back to the research lab several times.

People over 40 were excluded from the pool of subjects so as not to let presbyopia affect measurements of accommodative response. Subjects with no binocular vision were also excluded thanks to a simple Titmus stereoscopic test, in which subjects were asked to grab the wing of a fly presented in S3D.

Pre- and post-experimental measurements

After being accepted in the experiment, subjects were submitted to an array of measurements, so as to assess accommodative response and binocular status. Some were repeated after the experiment, to the goal of detecting significant changes in said measurements. Subjects with refractive anomalies were allowed to wear their usual correction (glasses or contacts).

Accommodative response

Accommodative response was evaluated pre- and post-experiment thanks to a lens flipper (+2.00 D / -2.00 D) test at 40 cm. Subjects were instructed to focus on a printed text through the lenses and signal the experimenter as soon as said text was in focus. The flipper was then rotated and the number of complete cycles per minute was recorded. Punctum proximum was also measured before and after the experiment, thanks to the Donder's push-up test: subjects were told to focus on a printed text placed at about 40 cm from their eyes. The text was then gradually moved closer and the distance at which it was impossible for them to focus on the text was recorded.

Stereoacuity

Stereoacuity was measured pre- and post-experiment, using the Titmus test circles, which exhibit decreasing disparities ranging from 800 sec arc to 40 sec arc.

Inter-pupillary distance

Inter-pupillary distance to infinity was measured with a pupillometer.

Convergence fusion and break points

Convergence fusion and break points were evaluated before and after the experiment. Since our lab did not have access to a haploscope, we devised a 'virtual haploscope', which consisted in the picture of an asterisk (a 5-point star with square apices) being projected onto the screen at varying disparities. Subjects were instructed not to move their heads for the perception of the projected images to remain unchanged except for their disparity, which varied. If a subject's fusion point was evaluated, the two images of the asterisk were projected with an excessive horizontal onscreen disparity of 40 cm (roughly equivalent to 15°, crossed), which gradually decreased until it reached zero. The subject was instructed to press a button when they could fuse both images together. To assess the subject's break point, both images of the asterisk were projected with a null disparity, which then gradually increased. The subject was instructed to press the button as soon as they experienced diplopia. The order in which both measurements were taken was randomized.

Binocular status

Following Lambooij et al. [25], the Wilkins Rate of Reading Test was administered to subjects before the experiment, in both 2D and S3D. The ratio between correctly read words in 2D and S3D was computed.

Subjective assessment of visual discomfort

After the experiment, subjects were asked to assess their state of visual discomfort through a questionnaire, modelled on Zeri and Livi's source factors [26]. Subjects were asked to assess each item (out of a total of ten) with a continuous rating scale, ranging from "Not at all" to "Absolutely".

Tasks

We designed four different tasks aimed at assessing distance and curvature perception on the Z-axis (depth axis). All of our experimental tasks were devised so as to potentially maximize asthenopia and visual discomfort within a short experiment time (the scenes are sharply focused throughout, the textures used to skin objects and backgrounds all contain extremely high spatial frequencies, tasks with depth movement involve changes from crossed to uncrossed disparity and vice-versa, etc.) Stimuli viewed in the context of S3D use at work would probably be less stressful to the visual system but exposure time would certainly be much longer than a mere 30 to 45 minutes, hence our decision to try to induce as much asthenopia and discomfort as possible in such a short time.

Each task is repeated 45 times before moving on to the next one. In between each occurrence, a one-second blank screen is shown. The results of the first five occurrences are discarded, as they are considered 'learning' occurrences. The four tasks are:

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

Positioning task

A ball travels on the Z-axis, passing below an arch. Subjects are instructed to press a button when the ball is exactly below the arch. The position of the arch and the starting position of the ball are randomized each time. The ball distance from the center of the arch on the Z-axis is recorded.

Depth perception task

Five spheres are interspersed with six cubes to form a line. Each sphere and cube is assigned a random depth. Subjects are asked to determine which sphere is closest to them, the cubes acting merely as distractors. The choice is made through pointing at the chosen sphere through hand tracking and then pressing a button. The Z error is computed and recorded. Response times are also logged.

Curvature perception task

A cylinder is shown with Z- and Y-scales (depth and height) randomly determined. Its horizontal position varies at each occurrence and is randomly picked between three possible positions (middle, left and right). Subjects are asked to appreciate whether it is flatter or more curved than normal and press buttons accordingly. Subjects' answers (flatter / curvier) are recorded, which allow for the computation, for each condition, of the point of subjective equality (PSE) and the just noticeable difference (JND). Response times are also logged.

Collision detection task

Two cubes of random scales are positioned on opposite 'walls'. They travel toward a random point on the opposite wall. Right before the cubes would overlap each other on the display, they disappear. Subjects are asked to appreciate whether the cubes would have collided had they not disappeared and press buttons accordingly. Correct / incorrect answers are logged, which allow for the computation of an error percentage. Response times are also recorded.

Experimental conditions

In the course of other research, we had observed that we could disable stereoscopy gradually, eventually showing a monoscopic image to subjects viewing supposed S3D contents, without their being aware of any change. As stated before (cf. Related work), stereoscopy, while being potentially stressful to the visual system, has been proved to aid perception in the case of near work, above all when coupled to motion parallax. We therefore decided to explore the idea of intermittent stereoscopy, i.e. an alternation of stereoscopic and monoscopic stimuli with gradual, linear transitions in between. The order in which both stimuli would appear brought us to devise two different experimental conditions:

- stereo stimulus at the onset of projection, linearly transitioning to monoscopic over 3 seconds —henceforth referred to as "Stereo @ beginning";
- monoscopic stimulus lasting for 4 seconds at projection onset, linearly transitioning to stereoscopic over 3 seconds — henceforth abbreviated "Stereo @ end".

These two experimental conditions were compared to two control situations:

- a monoscopic setup —referred to as "Mono"— in which the images presented to each eye were the same;
- an orthostereoscopic setup —henceforth called "Stereo"— which presented to both eyes images perceived by a parallel camera setup, the distance between both cameras being equal to the subject's inter-pupillary distance. It is also worthy of notice that all other stereoscopic stimuli presented to the subjects (which includes those presented in preliminary measurements —Rate of Reading Test in S3D— as well as those in intermittent stereo conditions) were likewise orthostereoscopic, not to overexert the subjects' visual systems with unnatural disparity.

The aforementioned conditions were randomized, thus leading to a double-blind experimental setup. The experimenter could however deduce which condition was being tested by merely looking at the screen without stereo goggles, and thus refrained from speaking or commenting, except to give instructions or answer subjects' questions. For subjects to avoid deducing which experimental condition was in progress, they were instructed to keep the goggles on at all times while the experiment was in progress. Head-tracking was enabled for all four conditions.

Experimental setup and apparatus

The experiment was carried out on a 4 x 3m screen on which images were retroprojected; two projectors were used to achieve a passive stereoscopy setup. Our lab used Infitec technology (DSPs and goggles) to present a different image to each of the subjects' eyes. Subjects' head and hand movements were tracked thanks to two ART tracking cameras. Subjects interacted with the program running the experiment through a flystick with various buttons they were to press. The protocol had been coded in C# inside of Unity3D and interfaced with our immersive setup thanks to MiddleVR.

Research hypotheses

Our experiment aimed at validating the following hypotheses:

- either or both condition(s) involving intermittent stereoscopy (Stereo @ beginning / Stereo @ end conditions) lead(s) to correct depth decisions or, failing that, to decisions as accurate as —or only marginally less accurate than— a stereo viewing context;
- any condition involving disparity cues (including intermittent stereo) yields a better precision than a mono setup;
- one or both of the intermittent stereoscopy conditions occasion(s) less visual fatigue and discomfort than classic S3D.

Results

Perception

Positioning task

We assessed response accuracy with the ball Z (depth) distance from the center of the arch in each test occurrence (the closer to 0, the more accurate), while the variance of said distance was considered a good indication of response precision (the smaller, the more precise). Figure 1 shows a graphical representation of the means of those distances with standard errors.

A one-way ANOVA was performed and was statistically significant [F (3, 2233) = 9.273, $p \approx 0$]. It was followed by Tukey's HSD post-hoc test, which indicated that the only non-significant pairwise comparisons were between Stereo @ end-Mono and Stereo @ beginning-Stereo. All other pairwise comparisons were statistically significant (p < .03). Brown-Forsythe tests were performed pairwise to determine whether variances were equal. The variances of all the conditions involving S3D significantly differed from that of Mono [Mono-Stereo: F (1,

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

1116) = 145.51, $p \approx 0$; Mono-Stereo @ beginning: F (1, 1076) = 6.200, $p = .012$; Mono-Stereo @ end: F (1, 1075) = 72.188, $p \approx 0$]. One can see from the graph that they are smaller than Mono's, indicating greater precision.

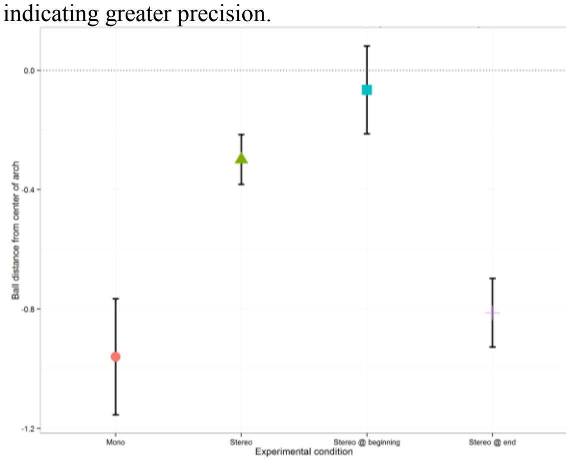

Figure 1. Task 1, Ball distance from center of arch: Means (with standard error)

Depth perception task

Response accuracy was assessed thanks to the Z error in subjects' answers (obtained through the formula: Z error = Z $\text{position}_{\text{correct answer}}$ – Z $\text{position}_{\text{chosen sphere}}$; the closer to 0, the more accurate). Said error's variance indicated response precision (the smaller, the more precise). Figure 2 shows a graphical representation of the data for Task 2.

A one-way ANOVA was performed and was statistically significant [F (3, 2396) = 4.738, $p = .003$]. It was followed by Tukey's HSD post-hoc test which indicated that subjects were much less prone to errors in the Stereo than in the Mono condition ($p = .002$). More interestingly, the Z error was also significantly less high in the Stereo @ beginning condition compared to the Mono condition ($p = .04$). All other pairwise comparisons were not statistically significant ($p > .05$). Pairwise Brown-Forsythe tests were performed to test for equality of variances: all conditions involving S3D significantly differed from Mono [Mono-Stereo: F (1, 1238) = 8.935, $p = .003$; Mono-Stereo @ beginning: F (1, 1158) = 8.354, $p = .004$; Mono-Stereo @ end: F (1, 1198) = 4.540, $p = .033$]. It is easily seen, from the graph, that they are smaller than Mono's, indicating greater precision.

Response times means and standard errors are also plotted in Figure 2. A one-way ANOVA was performed and was highly statistically significant [F (3, 2396) = 24.668, $p \approx 0$]. It was followed by Tukey's HSD post-hoc test which indicated that the response times in Stereo and Stereo @ beginning were very similar ($p = .7$). All other pairwise comparisons led to statistically significant differences between means of response times ($p < .003$)

Curvature perception task

As the experiment required subjects to answer either "flatter" or "curvier" than normal, Task 3 involved a forced choice paradigm. As a result, the data we obtained was processed through a psychometric function: we assumed a phi-gamma hypothesis and therefore fitted a cumulative normal distribution curve to the data. The Point of Subjective Equality (PSE) and the Just Noticeable Difference (JND) were computed, using probability values of .5 for the PSE and .8143 for the Differential Threshold (DT); the JND was calculated according to the formula: JND = DT – PSE. Although .8143 differs from the traditional .75 probability used to compute the DT, we chose to follow Bonnet [29] in considering —in the context of a phi-gamma hypothesis— that the JND should be defined as the PSE's standard deviation. Figure 3 shows the plotted experiment results, the fitted curves and the PSE and JND values. Response accuracy was assessed through PSE values (the closer to 1, the more accurate), when response precision was evaluated thanks to JND values (the smaller, the more precise).

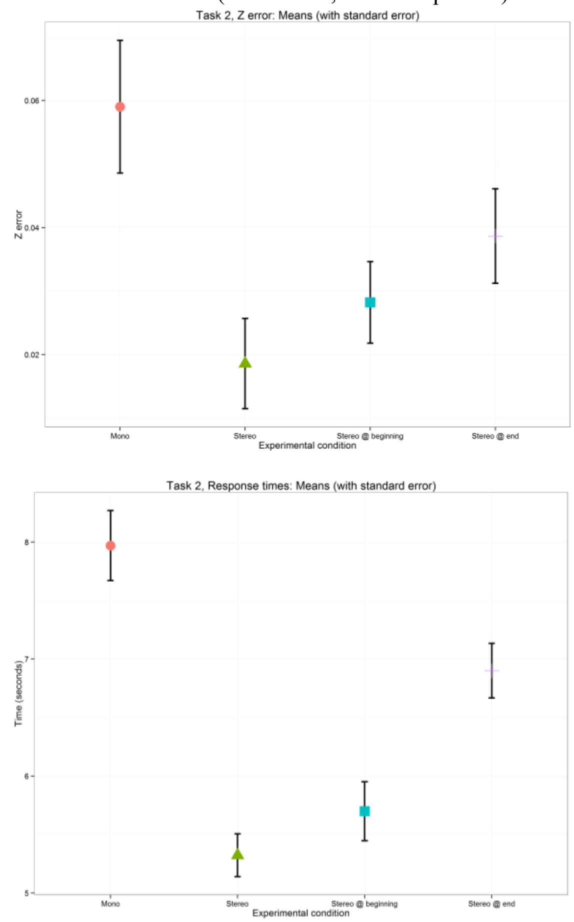

Figure 2. Task 2, Z error and Response times

A one-way ANOVA was performed and was statistically significant [F (3, 2098) = 24.652, $p \approx 0$]. It was followed by Tukey's HSD post-hoc test for pairwise comparisons: all pairs were deemed significantly different ($p < .01$), except for Mono-Stereo @ beginning, and Stereo @ end-Stereo ($p > .05$). Bartlett tests were performed pairwise to determine equality of variances. All variances were deemed significantly different ($p \approx 0$). The conditions leading to precise estimations are Stereo and Stereo @ end (although the former is significantly more precise than the latter). Stereo @ beginning has the worst precision of all.

Figure 4 shows the response time means and standard error. A one-way ANOVA yielded highly statistically significant results [F (3, 2393) = 25.334, $p \approx 0$], and was followed by Tukey's HSD post-hoc test. All pairwise comparisons were deemed statistically significant ($p < .01$) except for the pair Stereo @ beginning-Stereo @ end ($p > .7$)

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

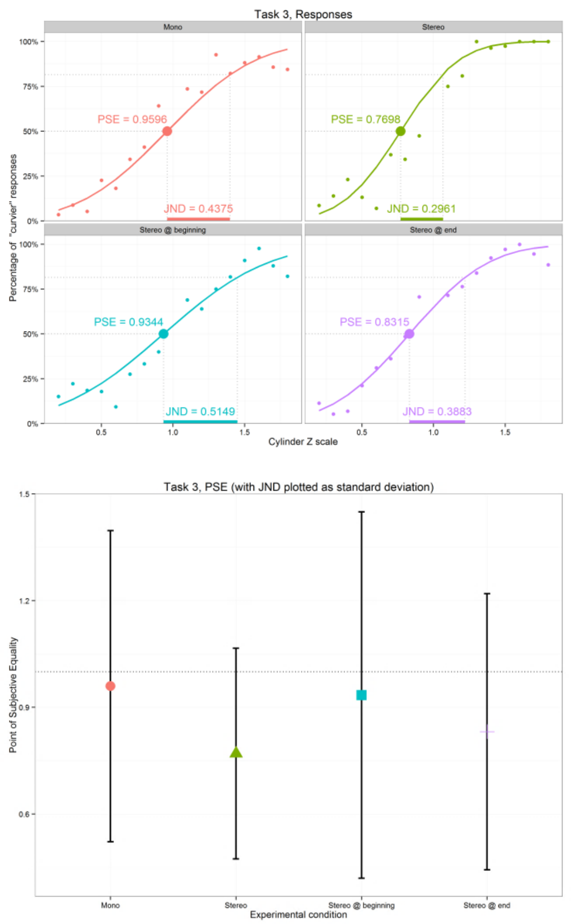

Figure 3. Task 3, Responses, and PSE values (with JNDs as standard deviation)

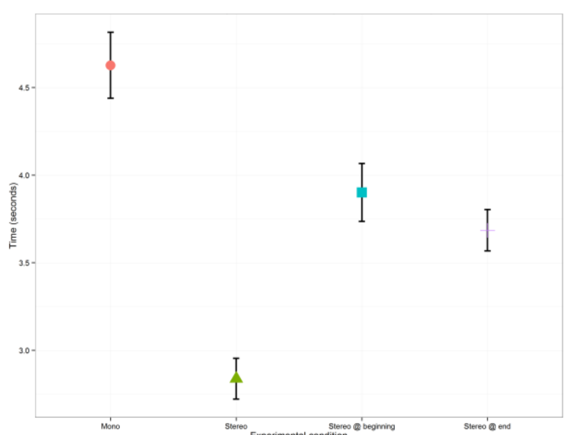

Figure 4. Task 3, Response times: Means (with standard error)

Collision detection task

We assessed response accuracy with the proportion of incorrect answers (the smaller, the more accurate). The percentages obtained, computed from a dichotomous categorical variable, were normalized with an arcsine transformation before being analyzed, according to the following formula: $x' = \arcsin(\sqrt{x})$. The variance in the proportion of errors hinted at response precision (the smaller, the more precise). Figure 5 shows the means of normalized errors per experimental condition, plotted with standard error.

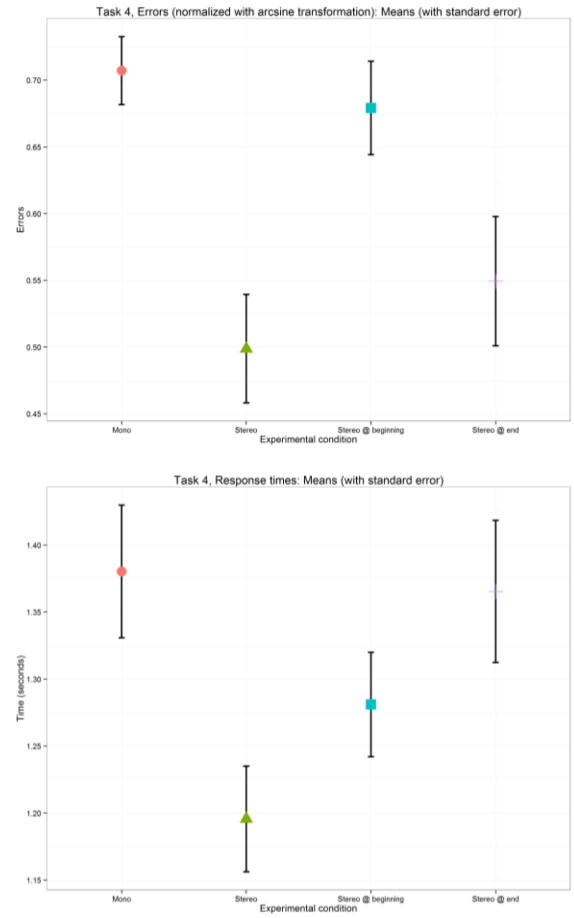

Figure 5. Task 4, Error percentages and Response times

We performed a one-way ANOVA on the normalized results, and found the means to be significantly different [F (3, 56) = 6.892, $p < 0.001$]. Tukey's HSD post-hoc test revealed, that three pairs had rather dissimilar means: Stereo-Mono (p = .002), Stereo @ end-Mono (p = .03) and Stereo @ beginning-Stereo (p = .01). Brown-Forsythe tests were performed pairwise to test variance equality: conditions involving S3D do not lead to increased precision compared to Mono, as the difference in variances is not statistically significant ($p > .2$).

The means of response times (with standard error) are also visible in Figure 5 and follow the same configuration as that observed for Task 2. We performed a one-way ANOVA on the response times which yielded statistically significant results, confirming the plot results [F (3, 2375) = 3.615, p = 0.013]. We then processed the results with Tukey's HSD post-hoc test, which revealed a significant difference between Stereo-Mono conditions (p = .01) as well as Stereo @ end-Stereo (p = .01). All other pairwise comparisons were statistically non-significant ($p > .4$).

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-041

Fatigue and discomfort

Optometric variables

For each variable in each of the conditions, we performed a one-tailed paired T-test on pre- and post-test measurements, following the alternative hypothesis of a difference in means smaller than 0 (i.e. each variable has decreased). Additionally, we performed, for each variable, a one-way ANOVA on the pre-/post-test difference.

Punctum proximum and ease of accommodation did not yield any statistically significant results. Additionally, the results for convergence fusion point were discarded as, to our surprise, numerous subjects were able to instantly fuse the high-disparity stimulus presented to them at the beginning of the evaluation of their fusion point from diplopia, thus experiencing no diplopia at all. This resulted in a truncated dataset.

In view of the very high *p*-values given by some of the T-tests for stereoacuity, we decided to perform a one-tailed paired T-test following the reverse alternative hypothesis (testing whether stereoacuity had increased after the experiment), which yielded a *p*-value of .045, which is significant. This isolated result could suggest a possible adaptation of the visual system to S3D, or simply be a false positive (the *p*-value for Mono was .065, which is non-significant, but not very different). This would have to be investigated much further, starting with a more reliable test (such as a random dot stereogram), as the Titmus test is well-known to be easily "cheated" because of the presence of monocular cues [30] [31].

As to the subjects' break points, one-tailed paired T-tests indicated that they had been affected by the Stereo condition [$t\,(11) = -2.132$; $p = .028$] and, to a lesser extent, by Stereo @ beginning (for which the difference in means was only marginally significant [$t\ (11) = -1.785$; $p = .051$]). The means in Mono ($p > .09$) and Stereo @ end ($p \approx .4$) were not significantly different. The fact that subjects' break points are affected by classic S3D is hardly remarkable; it is worthier of notice, however, that Stereo @ beginning may have a likewise effect. Even though the ANOVA we performed on the pre-/post- test differences was not statistically significant ($p = 0.282$), it is still interesting to take a look at the means of said differences (plotted in Figure 6). Stereo @ end seems, in our case, to have displaced subjects' break points the least. It did so much more consistently than other conditions: a couple of pairwise Brown-Forsythe tests performed on the variances of break point measurements differences (pre- / post-experiment) in Stereo @ end-Stereo [$F\,(1, 25) = 3.604$, $p = .069$] and Stereo @ end-Stereo @ beginning [$F\,(1, 25) = 7.623$, $p = .011$] confirmed that Stereo @ end performs significantly more consistently than any other S3D condition at not affecting subjects' break points.

Visual discomfort

A one-way ANOVA was performed on each of the 10 items present in the post-test subjective questionnaire, as well as another one-way ANOVA, this time on a global discomfort score (obtained, for each subject, by summing the scores they had given to each item in the questionnaire. They were all statistically non-significant.

Rate of Reading Test 2D/3D ratio

We performed a linear correlation test to emulate Lambooij et al.'s results stating that the 2D/S3D Rate of Reading ratio was a good predictor of binocular status, and therefore of visual discomfort25. The results we obtained [$r\,(58) = .256$, $p = .049$] did not permit us to replicate their findings.

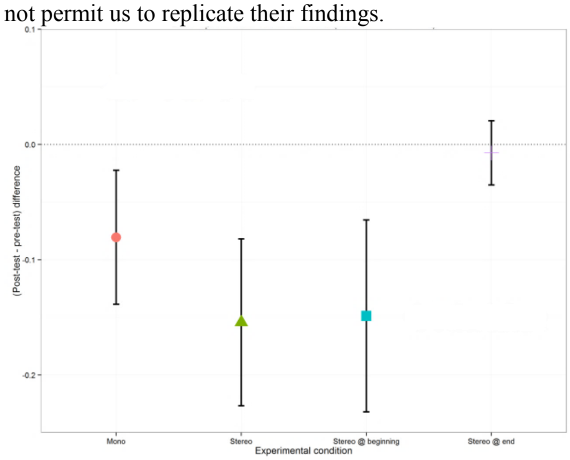

Figure 6. Convergence break point from fusion: Means (with standard error)

Discussion

Perception

Several assumptions can be made in the light of the results stated above.

Accuracy

We posit that the Stereo @ beginning condition leads to more accurate depth decisions than Mono when dealing with tasks requiring simpler mental operations, such as tasks with slow linear movement on the Z axis (Task 1) or discrimination of identical objects along a single axis with no movement (Task 2). This would suggest an ability in subjects to construct an initial depth map of the environment which the brain could keep using as long as modifications in said environment were not too dramatic. Task 3 was more problematic, because even though it exposed subjects to a simple environment (a static cylinder on a static background), it seems to us that assessing a curvature requires a much more complex mental process, leading to more mixed results for Stereo @ beginning (subjects in that condition do no better than Mono subjects). In a task like Task 4, with multiple objects (two cubes) and multiple movements on multiple axes (X and Z), it seems that S3D is needed when the depth decision-making process is to take place, thus giving a clear advantage to the Stereo @ end condition.

Precision

It seems that any condition involving S3D tends to lead to much better precision than monoscopic stimuli, then again only in the context of simpler environments and mental processes.

Response time

Stereo and Stereo @ beginning result in faster depth decision than either Mono or Stereo @ end. The presence of disparity stimuli at task onset seems to boost confidence in the initial depth judgement, decreasing response time.

Fatigue and discomfort

Even though the results we have presented above are not highly significant, and we cannot therefore wholly assert that intermittent stereoscopy occasions less visual fatigue and discomfort than full S3D, our data tentatively suggests that Stereo @ end may be less fatiguing for the visual system than Stereo @ beginning. We posit that this is due to the longer exposure time to S3D stimuli the latter condition occasions. Indeed, plotting response times for Tasks 2 and 3 —which are the only tasks in which there is no motion and subjects can therefore answer when they choose to— and superimposing the times when the perceived stimulus is actually S3D (as done in Figure 7) makes it clear that the Stereo @ beginning condition exposes subjects to S3D in a much greater proportion: indeed, subjects from the Stereo @ beginning condition responded 63% of the time when viewing stereoscopic stimuli, while subjects from the Stereo @ end condition were exposed to S3D during only 49.5% of their responses.

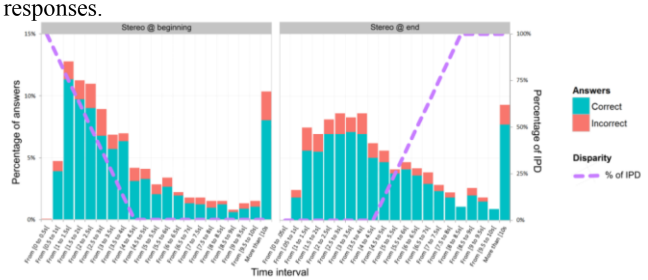

Figure 7. Tasks 2 and 3: Response times and disparity

This would therefore point to the fact that asthenopia, in intermittent stereoscopy, is merely proportional to the amount of exposure to S3D stimuli (even though calling it asthenopia is, in itself, another assumption, as the alteration of optometric variables has not been consistently correlated with subjective discomfort in the case of our experiment). In such a case, limiting asthenopia in such a viewing context would only consist in limiting subject exposure to S3D, entailing a judicious choice of when to present such stimuli.

Conclusions and future work

This paper tried to investigate the effects of intermittent S3D stimuli on perception and asthenopia. It seems S3D at task onset speeds up the depth decision-making process, and is sufficient for tasks involving a simple environment with little modification over time. However, the more complex the task becomes (mental operation, number of objects, multiple movements on several axes), the more S3D is needed when the decision is to be made. Additionally, working in S3D (be it constant or intermittent) is very likely to boost depth judgement precision, still in the case of simple environments or mental processes. It also seems that asthenopia experienced in intermittent stereoscopy viewing contexts is proportional to the amount of exposure to S3D stimuli.

Intermittent stereoscopy is, as of yet, vastly unexplored and this study is, to our knowledge, the first to investigate its effects. A huge amount of work remains to be done to reach a better understanding of how it works, and our tentative results need to be confirmed by other studies and experiments. Following are some pointers and possibilities for further studies on the matter:

- Assessing visual discomfort: it is extremely difficult to get a good estimate of subjective discomfort between subjects. A subsequent study would gain from adopting a within-subject experimental design, even if it is rather impractical to demand of subjects that they come back multiple times on different days.
- Stereopsis isolation: the results obtained in the course of this study are rather pooled together (insofar as the difference between the four experimental conditions is sometimes not clear-cut and definitive). A further study would maybe benefit from the suppression —to a feasible extent— of other depth cues and the isolation of stereopsis (as was done before4 5), so as to be able to differentiate the effects of intermittent stereoscopy from those of motion parallax, for example.
- Depth map lifetime: if the brain is indeed able to construct an environment depth map at task onset and keeps using it (even when stereoscopic stimuli are removed) in the case of simple mental operations, it would be of high interest to assess the time this environment model can last, and which factors occasion model obsolescence and therefore trigger depth re-mapping (presence of a new object, modification in movement, etc.)

References

[1] J. E. Cutting and P. M. Vishton, Handbook of perception and cognition, Vol 5; Perception of space and motion, San Diego CA: Academic Press, 1995.

[2] R. Patterson, “Human Factors of Stereoscopic Displays,” SID Symposium Digest Technical Papers, vol. 40, no. 1, pp. 805-807, 2009.

[3] M. T. Lambooij, M. Fortuin, W. A. IJsselsteijn and I. Heynderickx, “Measuring Visual Discomfort associated with 3D Displays,” Stereoscopic Displays and Applications XX, San Jose, California, 2009.

[4] M. Poyade, A. Reyes-Lecuona and R. Viciana-Abad, New Trends on Human–Computer Interaction, London: Springer-Verlag, 2009.

[5] J. Rivett and N. Holliman, “Stereoscopic Game Design and Evaluation,” Stereoscopic Displays and Applications XXIV, San Francisco, California, 2013.

[6] T. Grossman and R. Balakrishnan, “An Evaluation of Depth Perception on Volumetric Displays,” Working Conference on Advanced Visual Interfaces, Venezia, Italy, 2006.

[7] R. Smith, A. Day, T. Rockall, K. Ballard, M. Bailey and I. Jourdan, “Advanced stereoscopic projection technology significantly improves novice performance of minimally invasive surgical skills,” Surgical Endoscopy, vol. 26, no. 6, pp. 1522-1527, 2012.

[8] K. Votanopoulos, F. C. Brunicardi, J. Thornby and C. F. Bellows, “Impact of Three-Dimensional Vision in Laparoscopic Training,” World Journal of Surgery, vol. 32, no. 1, pp. 110-118, 2008.

[9] L. Leroy, P. Fuchs, A. Paljic and G. Moreau, “Study of stereoscopic shape perception for virtual design,” International Conference on Artificial Reality and Telexistence, Yokohama, Japan, 2008.

[10] L. Leroy, P. Fuchs, A. Paljic and G. Moreau, “Experiments on shape perception in stereoscopic displays,” Stereoscopic Displays and Applications XX, San Jose, California, 2009.

[11] E. D. Ragan, R. Kopper, P. Schuchardt and D. A. Bowman, “Studying the Effects of Stereo, Head Tracking, and Field of Regard on a Small-Scale Spatial Judgment Task,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 5, pp. 886-896, 2013.

[12] V. Cornilleau-Pérès and J. Droulez, “Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure,” Perception & Psychophysics, vol. 54, no. 2, pp. 223-239, 1993.

[13] M. T. Lambooij, W. A. IJsselsteijn and I. Heynderickx, “Visual Discomfort in Stereoscopic Displays: A Review,” Stereoscopic Displays and Virtual Reality Systems XIV, San Jose, 2007.

[14] T. Shibata, K. Joohwan, D. M. Hoffman and M. S. Banks, “The zone of comfort: Predicting visual discomfort with stereo displays,” Journal of Vision, vol. 11, no. 8, pp. 1-29, 2011.

[15] M. Wöpking, “Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus,” Journal of the Society for Information Display vol. 3, no. 3, pp. 101-103, 1995.

[16] Y. Okada, K. Ukai, J. S. Wolffsohn, B. Gilmartin, A. Iijima and T. Bando, “Target spatial frequency determines the response to conflicting defocus- and convergence-driven accommodative stimuli,” Vision Research vol. 46, no. 4, pp. 475-484, 2006.

[17] F. Speranza, W. J. Tam, R. Renaud and N. Hur, “Effect of Disparity and Motion on Visual Comfort of Stereoscopic Images,” Stereoscopic Displays and Virtual Reality Systems XIII, San Jose, 2006.

[18] R. S. Allison, “Analysis of the Influence of Vertical Disparities Arising in Toed-in Stereoscopic Cameras,” Journal of Imaging Science and Technology vol. 51, no. 4, pp. 317-327, 2007.

[19] J. C. Read and I. Bohr, “User experience while viewing stereoscopic 3D television,” Ergonomics, vol. 57, no. 8, pp. 1140-1153, 2014.

[20] A. W. Bangor, Display Technology and Ambient Illumination Influences on Visual Fatigue at VDT Workstations, Blacksburg, VA: Virginia Polytechnic Institute and State University, 2000.

[21] P. Neveu, A.-E. Priot, J. Plantier and C. Roumes, “Short exposure to telestereoscope affects the oculomotor system,” Ophthalmic and Physiological Optics, vol. 30, no. 6, pp. 806-815, 2010.

[22] S. Yano, S. Ide, T. Mitsuhashi and H. Thwaites, “A study of visual fatigue and visual comfort for 3D HDTV/HDTV images,” Displays vol. 23, no. 4, pp. 191-201, 2002.

[23] M. F. Fortuin, M. T. Lambooij, W. A. IJsselsteijn, I. Heynderickx, D. F. Edgar and B. J. Evans, “An exploration of the initial effects of stereoscopic displays on optometric parameters,” Ophthalmic and Physiological Optics, vol. 31, no. 1, pp. 33-44, 2011.

[24] E. Karpicka and P. A. Howarth, “Heterophoria adaptation during the viewing of 3D stereoscopic stimuli,” Ophthalmic and Physiological Optics, vol. 33, no. 5, pp. 604-610, 2013.

[25] M. T. Lambooij, M. Fortuin, W. A. IJsselsteijn, B. J. Evans and I. Heynderickx, “Susceptibility to Visual Discomfort of 3-D Displays by Visual Performance Measures,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 12, pp. 1913-1923, 2011.

[26] F. Zeri and S. Livi, “Visual discomfort while watching stereoscopic three-dimensional movies at the cinema,” Ophthalmic and Physiological Optics vol. 35, no. 3, pp. 271–282, 2015.

[27] S.-n. Yang, T. Schlieski, B. Selmins, S. C. Cooper, R. A. Doherty, P. J. Corriveau and J. E. Sheedy, “Stereoscopic Viewing and Reported Perceived Immersion and Symptoms,” Optometry and Vision Science, vol. 89, no. 7, pp. 1068-1080, 2012.

[28] H. Treiblmaier and P. Filzmoser, “Benefits from using continuous rating scales in online survey research,” Proceedings of the International Conference on Information Systems, Shanghai, China, 2011.

[29] C. Bonnet, Manuel Pratique de Psychophysique, Paris: Armand-Colin, 1986.

[30] H. Morris, A. R. O'Connor, M. G. Stephenson, M. Mitchell, G. J. Price and S. Anderson, “Clinical assessment of stereopsis and its functional significance,” British and Irish Orthoptic Journal vol. 2, pp. 8-14, 2005.

[31] S. L. Fawcett, “An evaluation of the agreement between contour-based circles and random dot-based near stereoacuity tests,” Journal of American Association for Pediatric Ophthalmology and Strabismus vol. 9, no. 6, pp. 572-578, 2005.

Author Biography

Ari A. Bouaniche has just graduated as valedictorian from a M.Sc. in Digital Creation and Publishing with a focus on Virtual and Augmented Reality at Université Paris 8, Saint-Denis, France. He is a Ph.D. candidate for the upcoming year. Bouaniche’s research focuses as a main interest on the study of the effect of virtual and augmented reality devices on the human visual system in terms of perception and visual fatigue. Bouaniche has also been exploring how gestural interactions occasion musculoskeletal diseases, and has investigated into the automated assessment of said interactions, based on several workplace safety standards in current use.

Laure Leroy graduated as a mecatronics engineer from the Polytechnic School of the Free University in Brussels, specializing in virtual reality. She therefore went to complete a Ph.D. in stereoscopic interfaces at Mines ParisTech, within the Center for Robotics, followed by postdoctoral studies in cognitive rehabilitation with virtual reality. Leroy is now an Associate Professor at Université Paris 8, Saint-Denis, France, where she studies the reduction of sensory-motor conflicts in virtual reality, in visual as well as gestural interfaces.