Measuring Visual Discomfort associated with 3D Displays

M. Lambooij[a,b], M. Fortuin[a,c,d], W. A. IJsselsteijn[a], I. Heynderickx[b,e]
[a]Human-Technology Interaction Group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands;
[b]Group Visual Experiences, Philips Research Laboratories, Eindhoven, The Netherlands;
[c]Department of Optometry, Hogeschool Utrecht, University of Applied Science, Utrecht, The Netherlands;
[d]Henry Welcome Laboratories for Vision Sciences, Department of Optometry and Visual Science, City University, London, UK;
[e]Department Mediamatics - Human Computer Interaction, Electronic Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands;

ABSTRACT

Some people report visual discomfort when watching 3D displays. For both the objective measurement of visual fatigue and the subjective measurement of visual discomfort, we would like to arrive at general indicators that are easy to apply in perception experiments. Previous research yielded contradictory results concerning such indicators. We hypothesize two potential causes for this: 1) not all clinical tests are equally appropriate to evaluate the effect of stereoscopic viewing on visual fatigue, and 2) there is a natural variation in susceptibility to visual fatigue amongst people with normal vision.
To verify these hypotheses, we designed an experiment, consisting of two parts. Firstly, an optometric screening was used to differentiate participants in susceptibility to visual fatigue. Secondly, in a 2x2 within-subjects design (2D vs 3D and two-view vs nine-view display), a questionnaire and eight optometric tests (i.e. binocular acuity, fixation disparity with and without fusion lock, heterophoria, convergent and divergent fusion, vergence facility and accommodation response) were administered before and immediately after a reading task.
Results revealed that participants found to be more susceptible to visual fatigue during screening showed a clinically meaningful increase in fusion amplitude after having viewed 3D stimuli. Two questionnaire items (i.e., pain and irritation) were significantly affected by the participants' susceptibility, while two other items (i.e., double vision and sharpness) were scored differently between 2D and 3D for all participants. Our results suggest that a combination of fusion range measurements and self-report is appropriate for evaluating visual fatigue related to 3D displays.

Keywords: Visual discomfort, visual fatigue, stereoscopic displays, optometric measurement, self-report

1 INTRODUCTION

Three-dimensional television (3D-TV) is often mentioned as the next generation of innovative displays for families to enjoy in their living room. It boosts the viewing experience by rendering content in front of or behind the display plane. A slightly different perspective of a scene is directed to each eye of the viewer, which provides the viewer with binocular disparity and thus stereoscopic depth, that is, the perception of depth based on the difference between the two retinal images. For a successful market introduction of 3D-TV, both the quality and comfort of the viewing experience must be at least comparable to that of conventional 2D-TV. This requirement has not yet been accomplished. Some viewers of 3D-TV still perceive visual discomfort, which is one of the reasons market introduction of 3D-TV is slowed down. In-depth research of the concept "visual discomfort" is needed in order to understand the perceptual mechanisms behind it, as well as its operationalisation in display evaluation.
In the literature, "visual discomfort" is used interchangeably with "visual fatigue", yet here they are distinguished. Visual fatigue refers to a decrease in the performance of the visual system as a consequence of physiological strain or stress resulting from excessive exertion[1]. In theory, it can be quantified objectively. Visual discomfort is its subjective counterpart[1]. A change within the visual system itself does not necessarily indicate visual fatigue. The visual system has some degree of plasticity and is able to adapt to altered viewing conditions. Only changes that decrease the performance of the visual system or that are accompanied by the experience of visual discomfort can be referred to as visual fatigue. The occurrence of visual discomfort alone, however, can be reason enough for further research. Firstly, consumers will not purchase a display that induces visual discomfort, even if the visual discomfort is harmless in terms of visual fatigue. And secondly and more importantly, absence of visual fatigue related to short-term viewing (e.g., five minutes) might

Stereoscopic Displays and Applications XX, edited by Andrew J. Woods, Nicolas S. Holliman, John O. Merritt, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 7237, 72370K

still compromise the visual system when longer viewing durations are used (e.g., two hours). Many 3D applications in entertainment settings concern long-term usage, e.g., 3D movies or 3D games. As such, carefully conducted long-term evaluations will be necessary to ensure that prolonged stereoscopic viewing does not induce any adverse side-effects to the visual system.

To determine the degree of visual fatigue and visual discomfort in a sensitive, accurate, reliable and valid way, multiple indicators for each component may be relevant. Indicators for visual discomfort can be provided by validated optometric questionnaires or other self-report measurements. Indicators for visual fatigue can be provided by clinical optometric measurement methods, which tend to be relatively cheap, concise, non-intrusive and quantitative. Moreover, they can easily be administered to a large group of participants. The measurement methods (or indicators) for both visual fatigue and visual discomfort should be relatively easy to apply in evaluative settings. Apart from that, they should fulfil a number of additional requirements[1]. Firstly, in order to address the impact of binocular depth on the visual system, the indicators should be able to distinguish stereoscopic viewing conditions from conventional monocular viewing conditions. Secondly, measurements should be relatively fast as the recovery of the visual system is usually quite fast. This constrains the length of the test. The recovery trajectory of the eyes after prolonged exposure to a stressful stimulus is in itself indicative of its functional plasticity and the severity of the visual strain. Thus, multiple measurements at different post-stimulus intervals may be needed. Thirdly, the indicators should apply to different types of displays, e.g., autostereoscopic systems and systems based on polarised or shuttered glasses. And lastly, measurements themselves should not require too much visual effort or induce visual fatigue or visual discomfort on their own.

Ideally, we would like to arrive at general indicators of visual fatigue and visual discomfort that can be implemented easily. When a robust relationship is established between visual discomfort and visual fatigue indicators, one indicator may be used to substitute the other. This would allow for example, the evaluation of the visual system for large groups of participants with simple subjective questionnaires, or the use of relatively simple objective measurements with participants that have difficulties in filling in questionnaires.

2 PREVIOUS RESEARCH

Previous research already applied clinical optometric measurement methods, in combination with questionnaires, to determine the effect of stereoscopic devices on the visual system. The results, however, were contradictory. Peli (1998) compared monocular and stereoscopic head-mounted displays (HMD) and a regular CRT on potential harmful effects to our visual system[2]. In his study, the binocular disparity values did not exceed one degree, implying that he remained within the accepted zone of comfortable viewing[1,3-6]. He used a set of objective indicators (e.g. refraction, visual acuity, fixation disparity, heterophoria and fusion measurements) and a subjective questionnaire as a pre- and post-measurement. Although no objective indicator revealed a significant, or clinically meaningful effect on any of the displays, almost all items on the questionnaire indicated lower comfort scores for the stereoscopic HMD than for the other displays. Emoto, Nojiri and Okano (2004) evaluated changes within the visual system as a consequence of viewing still images for 60 minutes in monocular and stereoscopic mode[7]. Visual fatigue was evaluated using a pre- and post-measurement of the fusion amplitude and the Accommodative Convergence / Accommodation ratio (AC/A ratio), i.e., the change in vergence due to accommodation per change in accommodation. Visual discomfort was evaluated via a questionnaire and a free-form in which participants could give their comments. To determine possible short term after-effects, fusion amplitude was measured after five and ten minutes rest. The results were related to the ability of the participants to free-fuse, because this might indicate larger a fusion amplitude. No differences were found between the pre- and post-measurements of the AC/A ratio, but the fusion amplitude of participants that were not able to free-fuse, significantly decreased in the convergent direction after stereoscopic viewing. The questionnaire revealed that five of the twelve participants experienced more visual discomfort, one experienced less visual discomfort and for six participants visual discomfort was similar under both stereoscopic and monoscopic viewing conditions. The subjective ratings were not related to the ability to free-fuse. More recently, Emoto, Niida and Okana (2005) performed an experiment in which participants viewed films for almost one hour stereoscopically, monocularly, and in a simulated stereoscopic conditon[8]. The simulated stereoscopic condition consisted of viewing monocular content through prisms. Prisms change the vergence, i.e., dissociate the visual lines of the eyes, while keeping the accommodation constant. The strength of the prisms was set according to each participant's individual Percival's area of comfort, which describes the range of prism loads that does not induce any discomfort[2,9]. Visual fatigue was measured with a pre- and post-measurement of the fusion amplitude and the accommodation response, while visual discomfort was measured with a post-questionnaire. Only one discomfort item, namely 'servere eye fatigue', was incorporated in the paper, and indicated more discomfort when the prism load was varying or beyond Percival's area of comfort. Both the accommodation response and the fusion amplitude were affected significantly by conditions with varying disparity in (simulated) stereoscopic conditions,

whereas the fusion amplitude also decreased significantly with fixed prism loads beyond as well as within Percival's area of comfort. No visual discomfort was perceived in the latter, which suggests that these changes in fusion amplitude indicated functional adaptations to altered viewing situations.
These studies revealed little consensus both between, as well as within, the indicators of visual fatigue and visual discomfort. Also large individual differences were revealed. Next to the differences in set-up of the experiments in these studies, we hypothesize two potential causes for these contradictions. The first cause is the differences in the visual system between the participants. More specifically, approximately 20% of the population has some form of a binocular anomaly[10]. The consequent visual complaints, which do not have to be present in normal viewing situations, may become present or more severe in unnatural viewing situations, e.g., viewing stereoscopic content. The second cause is that not all clinical tests are equally appropriate to evaluate the effect of stereoscopic viewing on visual fatigue and visual discomfort. Since both visual fatigue and visual discomfort are expected to decrease rapidly after viewing stereoscopic content, it makes no sense to compare all indicators after a long-term 3D video and it is too time consuming to let each participant view one long-term video per indicator. As such, as a first step short-term stimuli that stress the visual system are used that are assumed to represent long-term stimuli in terms of visual discomfort and visual fatigue. The purpose of the current study is to find empirical validation for both of these causes. Hence, the current experiment is aimed at answering the following two research questions:

1. What are the most appropriate indicators for short-term stimuli, that are assumed to be appropriate for long-term 3D videos as well, to evaluate visual fatigue and visual discomfort as a result of stereoscopic viewing?

2. What is the effect of the natural variation in susceptibility to visual fatigue and visual discomfort amongst people with normal vision?

3 EXPERIMENT SET-UP

3.1 Design

The experimental design consisted of two steps: (1) an optometric screening (1 session), and (2) the measurement of visual fatigue and visual discomfort under different conditions (4 sessions). The optometric screening allowed differentiating the participants with respect to their susceptibility to visual fatigue based on the functionality of their visual system (Susceptibility). The four sessions of the actual experiment differed in Dimension (2D and 3D) and in Display (nine-view auto-stereoscopic LCD and two-view polarised CRT), thus arriving at a 2x2 within-subjects design; the order of conditions was randomized across participants. In each session visual fatigue and visual discomfort were evaluated as a pre- and post-measurement (Pre-post). Eight different optometric indicators were used for visual fatigue and one indicator, a questionnaire containing 15 subjective items, was used for visual discomfort. The screening session and each experimental session were scheduled on different days. In total, 48 participants completed the experiment with the two-view display; only 18 participants completed the experiment with the nine-view display, due to the fact that we had this display on loan for a limited time only.

3.2 Screening of participants

Prior to the experiment, an extensive optometric screening was carried out on 50 naive participants. This screening was performed for three reasons: (1) to exclude participants with eye diseases or binocular abnormalities, (2) to familiarize participants with the optometric tests, since some of the screening tests were also used in the actual experiment, and (3) to differentiate participants on their susceptibility to visual fatigue associated with stereoscopic content.
The indicators with their exclusion criteria are outlined in Table 1 and contained three subjective questionnaires (the first three tests in Table 1) and 15 objective indicators. The three questionnaires relate to visual discomfort perceived whilst reading (CISS[11]), to dry eye problems (DEQ[12]), and to overall visual functions in everyday life (VFQ[13]). The CISS questionnaire was also used in the actual experiment because it incorporates items that relate to visual discomfort in general[14] and items to visual discomfort that relate to reading tasks specifically.
For the objective indicators, a short explanation is given here, and for a more extended description the authors would like to refer to Evans (2002)[10].Visual acuity measures the ability to have a clear vision. Refractive error refers to the defocus of the eye. Stereopsis indicates the ability to perceive stereoscopic depth. Fixation disparity indicates small misalignments of the eyes that were measured with and without fusion. Heterophoria refers to the situation in which the visual lines of the two eyes do not intersect at the fixation point. The cover test is one specific measurement of heterophoria, i.e., by covering one of the two eyes. Convergent and divergent fusion indicate the amount of convergence

and divergence respectively that can be induced before fusion is compromised and double vision occurs. They are commonly characterized with prism loads at which binocular single vision is lost (break point) and recovered again (recovery point). The near point of convergence gives the shortest distance to which the eyes can converge, and the accommodation amplitude gives the shortest distance to which the eyes can focus. The accommodation response refers to the focus control of the eye, i.e., the accuracy of the accommodation in direction and size. Accommodation and vergence facility indicate the adaptive capacity of the accommodation and vergence systems measured with flippers, i.e., two pairs of lenses or prisms to which participants have to adjust whilst retaining clear binocular single vision. The number of flips participants can make per time unit indicates the facilities. The slitlamp microscope assesses the extent of dry eyes.

Table 1 The optometric indicators, both objective and subjective, that were applied in the screening including their measurements aspects and their exclusion criteria.

Indicators	Measurement aspects	Exclusion criteria
CISS	Convergence Insufficiency Symptom Score with 15 items	
DEQ	Dry Eye Questionnaire with 11 items	
VFQ	Visual Functioning Questionnaire with 25 items	
visual acuity	monocular and binocular	< 0.8 for both monocular and binocular
refractive error	monocular and between eyes	> 1 Diopter
stereopsis	Randot Stereotest	> 60"
fixation disparity	Mallet aligning prims with OXO fusion lock (near distance)	
heterophoria	Maddox rod (far distance)	vertical deviation > 1 PD
heterophoria	Maddox wing (near distance)	vertical deviation > 1PD
cover test	(near and/or far) heterophoria or heterotopy (strabismus)	if strabismus*
convergent fusion range	break-recovery method (far distance)	Sheard's criterion**
divergent fusion range	break-recovery method (far distance)	Sheard's criterion
near point of convergence		
accommodation amplitude	Binocular push up test	
accommodation response	MEM-retinoskopy	
accommodation facility	accommodation flipper binocular	
vergence facility	prism flipper binocular	
slitlamp microscope	fluorescein staining/ Break Up Time tearfilm / blepharitis	Grading: degree ≥ 2 (out of 4)

* strabismus is defined as a permanent deviation between the two eyes
** Sheard's criterion refers to a fusion range that still allows comfortable viewing

Based on the exclusion criteria, two of the 50 participants were excluded from the experiment, resulting in 48 participants with normal visual functions. Even though their visual system was characterized as normal, it still may differ in performance from participant to participant. The latter was investigated with a modified algorithm proposed by Evans (2002)[10] to evaluate the degree of decompensated heterophoria. Decompensated heterophoria refers to a small misalignment of the eyes that is not compensated by fusion mechanisms. Especially in unnatural viewing situations, e.g., viewing stereoscopic content, fusion mechanism may become inadequate to compensate heterophoria and people may experience visual complaints. The algorithm is outlined in Table 2. It computes a score representative for the degree of decompensated heterophoria. This score is the accumulative value of ten single scores that each relate to the result of an optometric measurement. The algorithm was modified by excluding indicator number 6 and 8. Foveal suppression (nr 6) concerns a test for near viewing and as such was not adequate for our experiment (the viewing distance in our experiment was three meters). Percival's criterion (nr 8) is defined as the middle third of the "zone of clear, single binocular vision"[9,10]. This zone can be determined by measuring the blur points of our fusion range, i.e., the points at which *clear* vision is lost. In the screening, however, the fusion range was determined by measuring the break points, i.e., the points at which binocular *single* vision is lost. Previous research has revealed that Percival's criterion is not suited for far distances[1]. The total algorithm score (excluding nr 6 and 8) ranged from 0 to 13; a score of 4 or lower could be labelled as less susceptible to visual fatigue[6]. When using this threshold, the algorithm divided the participants in 42 persons being less susceptible and 6 persons being susceptible to visual fatigue.

Table 2 Algorithm proposed by Evans (2002) to differentiate participants with respect to their visual health based on objective and subjective optometric indicators[10]. Indicator number 6 and 8 were excluded from the algorithm used in this paper.

Nr	Indicators	Sign or symptom	Score
1	decompensated heterophoria	One or more of the questionnaire symptoms (if so, +3 or +2 or +1 if borderline)	3
2	cover testing	heterophoria detected	1
3	cover testing	recovery rapid and smooth (if so, +2 or +1 if borderline)	2
4	fixation disparity	aligning prism (Mallet): <1 Δ patients under 40 or <2 Δ patients over 40	2
5	fixation disparity	aligning prism (Mallet): stable	1
6	*foveal suppression*	*foveal suppression (Mallet): >3', or diplopia*	*2*
7	Sheard's criterion	failed	2
8	*Percival's criterion*	*Percival's criterion*	*1*
9	dissociated heterophoria	unstable	1
10	fusion amplitude	< 20 Δ	1

3.3 Equipment

The measurements during the screening were performed on a work station that included a control console, an examination chair, a double sliding instrument table, a projector column and a phoropter arm. The phoropter arm contained prisms and lenses to evaluate the visual functions described in the screening. The program Test Chart 2000 was used to evaluate the visual functionality of the participants during screening and during the experiment. It included a range of vision assessment tools (e.g. visual acuity, fusion, phoria and fixation disparity tests) for far viewing distances displayed on a separate CRT monitor placed at the same distance as the stereoscopic displays during the experiment (three meters).

The stimuli for the actual experiment were displayed on two different stereoscopic displays: a 42" Philips autostereoscopic nine-view lenticular LCD display[15] and a 20" two-view CRT polaroid stereoscopic display with a half transparent mirror in between. The nine-view lenticular LCD display had a resolution of 1600 x 1200 and the successive nine-views had a total width of 21 degrees. The two-view CRT display had a resolution of 720 x 567 and polariod glasses were required to direct the correct view to the correct eye. Note that by incorporating both displays we tend to remain display-independent in stead of comparing them when evaluating the effect of stereoscopic viewing on visual fatigue and visual discomfort.

3.4 Stimuli

The stimuli in the experiment were four different stereoscopic and monoscopic passages of the Wilkins Rate of Reading test[16] (WRRT). It consists of 10 lines with on each line the same 15 words distributed randomly (e.g., "you for the and not see my play come is look dog cat to up"). The text is independent of any syntactic and semantic constraints, i.e., participants did not know which words came next, which required them to remain focused on the text. A solid black frame was added to the text to allow faster and easier fusion.

The amount of screen disparity in the 3D condition depended on the display. On the nine-view display the maximum amount of screen disparity was set, which corresponded to a screen disparity of 0.4 degrees. With the two-view display a screen disparity of 1.5 degrees was set. The stimuli on both displays were quite stressful to be sure visual fatigue and visual discomfort were induced; the main stressor of the two-view display was the high amount of screen disparity and the main stressor of the nine-view was the high amount of crosstalk, which is perceived as blur[1,17,18].

3.5 Procedure

Participants were provided with an informed consent containing information about the screening and the experiment, and about the possible occurrence of unharmful visual discomfort. After signing the informed consent, they proceeded with the optometric screening. The tests applied in this screening are outlined in Table 1, and required about 45 minutes to complete.

Those participants that completed the screening successfully, participated in the experiment. They were seated at a viewing distance of three meters and received a brief instruction about the course of the experiment. All questions

concerning the procedure of the experiment were answered, after which the experiment started. Table 3 provides an overview of the procedure of the experiment. The first column outlines the order of the subjective and objective indicators. The participants began and ended with a stimulus followed by the CISS questionnaire for the pre- and post-evaluation of visual discomfort. In between eight objective indicators were measured, each before and after a stimulus. The stimulus was one of the four passages of the WRRT that was assigned randomly to each indicator. The participants were asked to read the text 'out loud' for 60 seconds. The indicators used are incorporated in the second column of Table 3. Their order is described in the third column of Table 3. A 'pseudo' randomization was applied to avoid visual fatigue induced by the tests themselves as much as possible. To this end, the objective indicators were divided into three blocks. Since the indicators in the first block did not require any visual effort, they were used to start with and were randomly administered. The order of the indicators in the second block depended on the participant's direction of the heterophoria, i.e., convergent or divergent. The participants first performed the fusion test in the direction opposite to their heterophoria (for compensation), and then the fusion test in the same direction as their heterophoria. The two indicators in the last block could require some visual effort, and therefore, were postponed to the end of the experiment. They were again mutually randomized. The fourth column of Table 3 provides the tests used to measure the indicators. For a description of these tests we refer to Evans (2002)[10]. The CISS questionnaire included the following items: uncomfortable, loss of concentration, double vision, sleepiness, sharpness, exhaustion, appearance of moving words, slower reading, loss of position in text, trouble remembering words, reread words, headache, pain in the eyes, strain in the eyes and irritated eyes. The participants performed the experiment four times: varying dimensionality (2D vs. 3D) and type of display (two-view vs. nine-view). Each session required about 45 minutes.

Table 3 Overview of the procedure of the experiment including the arrangement of objective and subjective indictors, the objective indicators and their order and the tests used to measure all indicators.

Test order	Indicators	Randomization	Tests	
Pre-subjective	-CISS (after first reading task)		-questionnaire	
Block 1 objective	-visual acuity binocular -fixation disparity (with OXO fusion lock) -fixation disparity (without fusion lock)	randomized within block 1	-Log Mar -aligning prism -red/green nonius lines	Test Chart 2000
Block 2 objective	-heterophoria -divergent fusion range -convergent fusion range	order based on direction of heterophoria	-Maddow rod -break / recovery by Risly rotary prism	Test Chart 2000
Block 3 objective	-vergence facility -accommodation response	randomized within block 1	-prism flipper -MEM-retinoskopy	
Post-subjective	-CISS (after last reading task)		-questionnaire	

4 RESULTS

An ANOVA was performed with Display, Dimension, Pre-post and Susceptibility as independent variables and the test results as dependent variable. The ANOVA revealed differences between the two displays ($p < 0.05$) and as such an ANOVA was performed per display.

4.1 Two-view display

Susceptibility to visual fatigue

Two indicators, fusion amplitude and vergence facility, were able to differentiate between susceptible and less susceptible participants, as outlined in Table 4. The number of flips participants could make in 60 seconds ($F(1, 164) = 12.832$, $p < .001$) as well as the break values of the total fusion amplitude ($F(1, 162) = 6.736$, $p < .01$) differed between susceptible and less susceptible participants significantly. The recovery value of the total fusion amplitude lacked significance ($F(1, 162) = 2.747$, $p < .099$). Participants that were susceptible to visual fatigue revealed clinically meaningful lower number of flips and break and recovery values than those that were less susceptible. Participants that were susceptible also indicated more visual discomfort than those who were less susceptible ($F(1, 24302) = 17.582$, $p < .001$).

Objective visual fatigue

When all participants were pooled together, none of the indicators was able to differentiate significantly between the 2D and 3D condition or between the pre- and post-measurement. When considering only those participants that were characterized as susceptible to visual fatigue, only in fusion amplitude a clinically meaningful increase in the post-measurement of the 3D stimulus was found, as outlined in Table 4. Both the break and recovery value of this group of participants showed an increase of 4 and 5 prism diopter (PD) respectively, between the pre- and post-measurement of the 3D stimulus. A separate analysis on the direction of the fusion amplitude, i.e., convergent or divergent, revealed that the fusion amplitude in convergent direction accounted for this increase. Figure 1 visualises this effect for the recovery value in the convergent direction. From the six participants that were characterized as susceptible, two were not able to fuse the 3D stimulus.

Table 4 Results of the objective indicators fusion amplitude and vergence facility for the two-view display.

Test		all participants (n = 48)				less susceptible (n = 42)				susceptible (n = 6)			
		pre		post		pre		post		pre		post	
		br*	rec*	br	rec	br	rec	br	rec	br	rec	br	rec
total fusion	2D	32	21	34	22	33	21	35	23	23	17	22	16
	3D	32	23	32	24	33	23	34	25	22	15	**26**	**20**
convergent fusion	2D	23	15	25	16	24	16	26	17	14	10	13	9
	3D	23	18	24	19	24	18	25	19	16	10	**19**	**15**
divergent fusion	2D	9	6	9	6	9	6	9	6	7	6	8	6
	3D	8	6	8	6	9	6	8	6	6	5	7	5
vergence facility	2D	10		11		11		12		5		6	
	3D	10		11		11		11		4		5	

* For each pre- and post-measurement of the fusion amplitude the first value denotes the break (br) value and the second value the recovery (rec) value. The bold values denote the increased fusion ranges of the participants in the susceptible subgroup that were able to fuse the 3D stimulus.

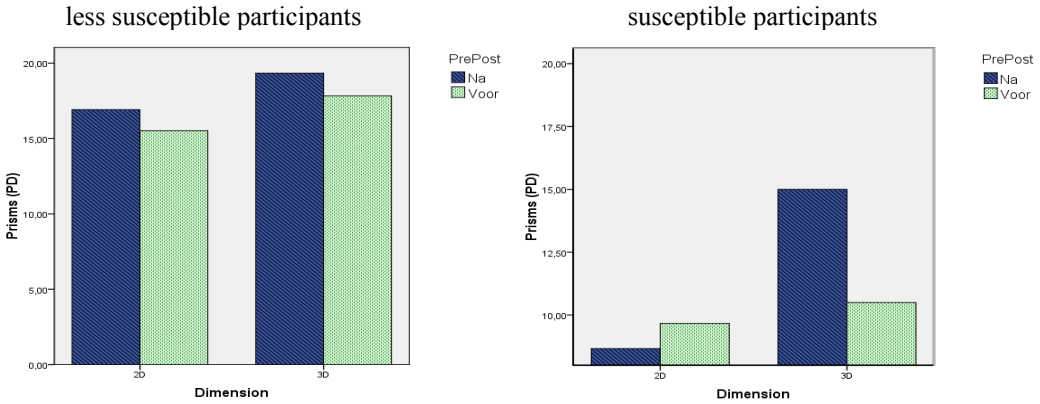

Figure 1 Mean recovery value of the fusion amplitude in convergent direction for the two-view display per degree of susceptibility. The x-axis represents the variation in dimension, the y-axis represents the averaged measurement values in prism diopter (PD) and the different lines represent the pre- and post-measurements.

Subjective visual discomfort

The grouping of all items of the questionnaire indicated more visual discomfort in the 3D than in the 2D condition ($F(1, 2430) = 161.466$, $p < .001$) and more in the post- than in the pre-measurement ($F(1, 2430) = 48.050$, $p < .001$). A few aspects are noteworthy. Firstly, most of the items yielded a low averaged visual discomfort score, referring to moderate

levels of visual discomfort. Secondly, of the three independent variables, Dimension seemed to have the largest effect on visual discomfort, whereas the effect of Pre-post on visual discomfort seemed smallest. And thirdly, not all items were affected equally by changes in the independent variables. This is reflected in a number of significant interaction effects; Item x Dimension ($F(14, 2430) = 6.493$, $p < .001$), Item x Susceptibility ($F(14, 2430) = 2.724$, $p < .01$), and Item x Pre-post ($F(14, 2430) = 1.222$, $p < .251$).

A principal component analysis[19] (PCA) was performed to find intrinsic correlations between the 15 items and to reveal if certain items shared similar underlying attributes of visual discomfort. The PCA was combined with an orthogonal rotation method (Varimax) to minimize the number of items with high factor loadings on more than one factor (i.e., a linear combination of the original 15 items). One of the assumptions of PCA, however, is that the data should be independent. Because a within-subject design was used, this assumption was violated. Therefore, the PCA was first performed per condition, i.e., 2D pre, 2D post, 3D pre and 3D post, as well as on the questionnaire results of the screening. Each PCA revealed almost identical results after the Varimax rotation with respect to the number of factors with high eigenvalues and the factor loadings of each item on these factors. As such, it was decided to perform a PCA on the entire data set. The resulting PCA revealed three underlying factors that explained 43%, 18% and 7% of the variance of the data. Factor 1 received high factor loadings of the items uncomfortable, double vision, moving words, slower reading, sharpness, loss of position in text and reread words. Factor 2 consisted of exhaustion, pain, irritation and strain. Factor 3 comprised of headache, sleepiness and trouble remembering words. Figure 2 depicts the factor loading of each item on the first two factors. A reliability test was performed that analyzed the internal consistency between the items belonging to a given factor. Based on this analysis, it was decided to exclude the item concentration from further analysis, since it was insufficiently consistent neither with factor 1 nor with factor 2. Reliability testing revealed Cronbach's alphas of 0.91, 0.85 and 0.67 for factor 1, 2 and 3 respectively.

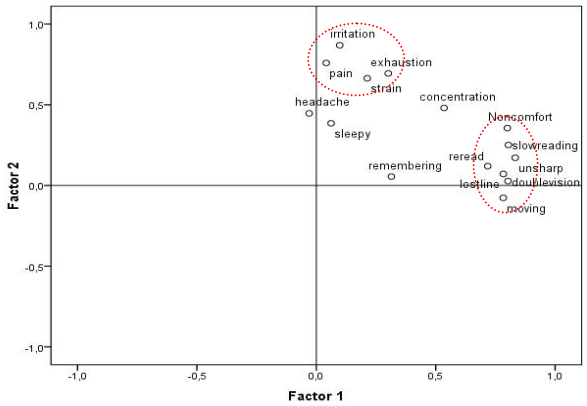

Figure 2 Factor loadings of all visual discomfort items on the two main factors for the two-view display. The circles indicate the items that were assigned to the two factors.

For each participant a factor score was calculated for each factor in each condition; i.e., the score given by the participant on each questionnaire item was weighted with the factor loading of that item, and then summed over all items of the factor. An ANOVA with Susceptibility, Dimension and Pre-post as independent variables and the factor scores as dependent variables revealed that factor 1 was only affected significantly by a change in Dimension ($F(1, 162) = 25.767$, $p < .001$), factor 2 was only affected significantly by Susceptibility ($F(1, 162) = 20.732$, $p < .001$) and factor 3 was not affected significantly by any of the independent variables. Items with high factor loadings on factor 1 and low factor loadings on factor 2 (i.e., double vision and sharpness) responded specifically on a change in Dimension. Similarly, items with high factor loadings on factor 2 and low factor loadings on factor 1 (i.e., pain and irritation) responded specifically on differentiation in susceptibility to visual fatigue. Figure 3 depicts the visual discomfort scores of these items as a function of Susceptibility and Dimension.

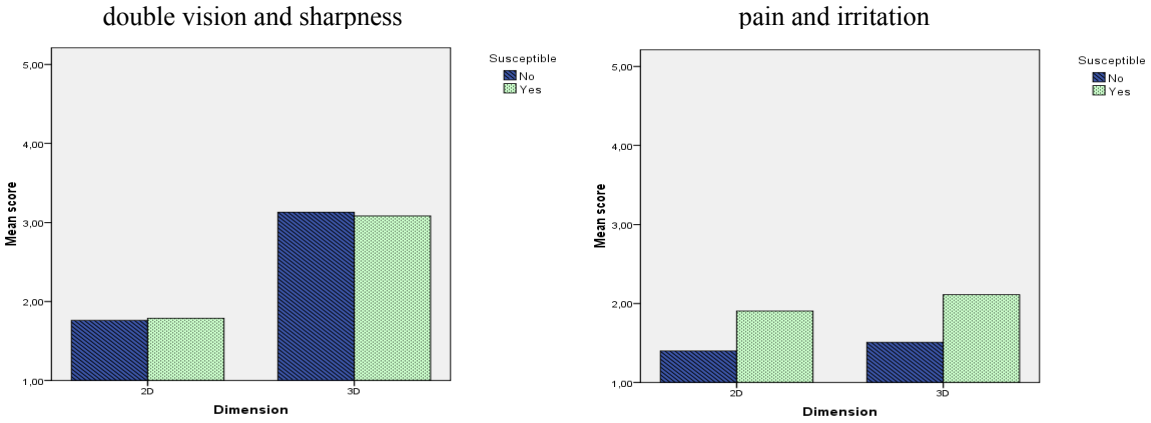

Figure 3 Average over the mean scores of the subjective items double vision and sharpness in the left graph, and pain and irritation in right graph for the two-view display. The x-axis represents the variation in dimension, the y-axis represents the averaged score and the different lines represent the degree of susceptibility.

4.2 Nine-view display

Susceptibility to visual fatigue

Similarly as with the two-view display, fusion amplitude and vergence facility, were able to differentiate between susceptible and less susceptible participants, as outlined in Table 5. Participants that were susceptible to visual fatigue revealed clinically meaningful lower number of flips and break and recovery values than those that were less susceptible. Participants that were susceptible also indicated more visual discomfort than those who were less susceptible, as depicted in Table 5. Though due to the small number of participants in this group (n=2), these differences between susceptible and less susceptible lacked significance.

Objective visual fatigue

When evaluating all participants as one group, none of the indicators was able to distinguish neither between 2D and 3D nor between the pre- and post-measurement significantly. In contrast to the two-view display, the group of participants being susceptible to visual fatigue did not reveal a significant difference between the pre- and post-measurement in the 3D condition. Noteworthy is that the convergent fusion values with the nine-view display were lower than with the two-view display.

Table 5 Results of the objective indicators fusion amplitude and prism flipper for the nine-view display.

Nine-view display															
		all participants (n = 18)				less susceptible (n = 16)				susceptible (n = 2)			
Test		pre		post		pre		post		pre		post	
		br	rec	br	rec	br	rec	br	rec	br	rec	br	rec
total fusion	2D	33	25	33	25	34	26	35	26	19	14	21	17
	3D	34	25	34	27	36	26	36	28	19	15	20	14
convergent fusion	2D	24	19	25	20	26	21	26	21	11	8	13	10
	3D	25	19	26	21	27	21	28	23	12	9	12	9
divergent fusion	2D	9	6	9	6	9	6	9	5	8	6	8	7
	3D	9	6	8	6	9	6	8	6	8	6	8	6
vergence facility	2D	14		15		15		16		6		8	
	3D	14		14		15		15		6		6	

Subjective visual discomfort

Overall, the items in the questionnaire indicated more visual discomfort in the 3D than in the 2D condition ($F(1, 1020) = 168.619$, $p < .001$) and more in the post- than the pre-measurement ($F(1, 1020) = 28.602$, $p < .001$). Again, almost all the items yielded a low averaged score, referring to moderate levels of visual discomfort. And not all items were affected equally by changes in the independent variables, which is reflected in one significant interaction effect; Item x Dimension ($F(14, 1020) = 9.823$, $p < .001$), Item x Susceptibility ($F(14, 1020) = 1.477$, $p < .112$) or Item x Pre-post ($F(14, 1020) = 1.366$, $p < .163$).

Following a similar analysis as performed for the two-view display, the PCA revealed four underlying factors that explained 38%, 14%, 12% and 7% of the variance in the data. Reliability statistics revealed Cronbach's alphas of 0.89, 0.86, 0.75 and 0.63 for each of the four factors, respectively. Especially factor 1 and 2 were of interest as they were relatively similar to factor 1 and 2 found in the analysis of the two-view display. Factor 1 received high factor loadings of the questionnaire items uncomfortable, double vision, moving words, and sharpness, while factor 2 received high factor loadings of the items exhaustion, pain, irritation and strain. Factor scores were calculated and an ANOVA with Susceptibility, Dimension and Pre-post as independent variables and the factor scores as dependent variables was performed. It revealed that factor 1 only responded significantly to a change in Dimension ($F(1, 68) = 30.881$, $p < .001$) and factor 2 responded significantly to Susceptibility ($F(1, 68) = 7.701$, $p < .01$) and Pre-post measurement ($F(1, 68) = 7.222$, $p < .01$). Factor 3 and 4 did not respond significantly to any of the independent variables. Similarly as for the two-view display, the items double vision and sharpness responded specifically to a change in Dimension. Though the items pain and irritation were strongly related to factor 2, they did not respond specifically to differentiation in susceptibility. Figure 4 depicts the visual discomfort scores of these items as a function of Susceptibility and Dimension.

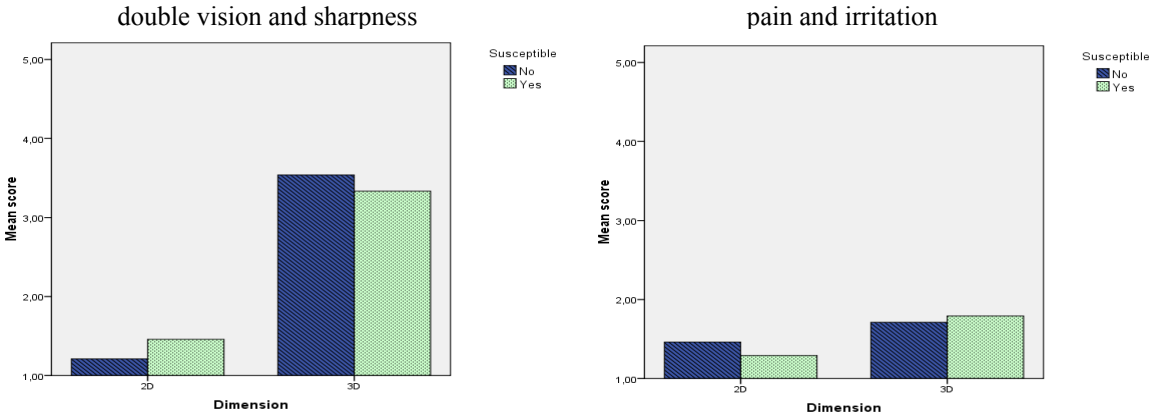

Figure 4 Average over the mean scores of the items double vision and sharpness in the left graph and pain and irritation in the right graph for the nine-view display. The x-axis represents the variation in dimension, the y-axis represents the averaged score and the different lines represent the degree of susceptibility.

5 DISCUSSION

We hypothesized two potential causes for the contradictory results in previous research on visual fatigue related to 3D displays: 1) not all clinical tests are equally relevant to evaluate the effect of stereoscopic viewing on visual fatigue and visual comfort, and 2) there is a natural variation in susceptibility to visual fatigue and visual discomfort amongst people with normal vision. Therefore, we performed an experiment, in which different clinical objective and subjective tests were compared under similar experimental conditions in order to find evidence for both potential causes.

With respect to the first hypothesis, most objective indicators did not reveal any change in visual functionality caused by 3D stimuli with neither of the displays. Under certain specific circumstances, however, some indicators did show clinically meaningful changes after the 3D stimuli accompanied by the experience of visual discomfort. Hence, these indicators are more sensitive to visual fatigue associated with stereoscopic displays. These specific circumstances underline the support of the second hypothesis, i.e., the fact that certain people respond differently to stereoscopic content as a result of differences in their visual system. Only people with relatively poor visual capacities revealed changes in fusion amplitude after the 3D stimuli. Hence, we can differentiate people's degree of susceptibility to visual fatigue associated with stereoscopic displays based on their visual system. We also found that a combination of fusion

range measurement and self-report is appropriate for measuring visual fatigue and visual discomfort associated with 3D displays. This is in line to some extent with previous research in which significant changes were found within fusion amplitude only in participants who were unable to free-fuse stereoscopic stimuli[7].

Although both hypotheses tested in this experiment are confirmed, some aspects need clarification. Firstly, the size of the increase in fusion amplitude related to the amount of disparity in the 3D stimuli. The 3D stimuli on the two-view display had much more disparity than on the nine-view display, which explains why only changes in fusion amplitude on the two-view display were found. The direction of the increase was similar to the direction of the 3D stimuli, i.e., only changes in fusion amplitude were found in convergent direction. In optometry, an increase in fusion amplitude indicates an improvement of the visual performance. Indeed, the participants seemed to adapt to a changed viewing situation, however, accompanied by the experience of visual discomfort. Even more, the change in fusion amplitude was only present within the group of participants that were susceptible to visual fatigue. As such, it is referred to as visual fatigue in stead of as a functional adaptation to a change in the environment or an improvement of the visual performance.

Secondly, one can wonder whether the indicators found for visual fatigue and visual discomfort for short-term stimuli are appropriate for long-term stimuli as well. Since only non-moving stimuli were used in this experiment, other indicators might be more relevant for stimuli with varying disparity, e.g., 3D films. Although this may be true, there is no reason to assume that fusion amplitude is a less relevant indicator for moving stimuli before studies reveal otherwise. Both moving as well as non-moving stereoscopic stimuli may impose strain on multiple properties of the visual system including fusion characteristics. And Emoto et al. (2005) already revealed that the fusion amplitude is an appropriate indicator for moving stimuli and stimuli with less disparity[4].

And finally, it should be noted that the relationship between objective indicators for visual fatigue and subjective indicators for visual discomfort can still not be accomplished. Changes within the visual system as a result of stressful short-term stereoscopic still images have a rapid deterioration, which complicates the detection of visual fatigue with pre- and post-measurement of optometric indicators. This rapid deterioration may have been the reason why Peli (1998) did not reveal any clinically meaningful visual fatigue, because he performed all his tests as a set before and after a stimulus[2]. On the other hand, people seem very capable in identifying the perceived visual discomfort. Participants experienced more visual discomfort in 3D than in 2D, more in the post- than in the pre-measurements, and participants that were susceptible experienced more visual discomfort than those who were less susceptible. Visual fatigue, however, was only measured after viewing 3D at participants who were susceptible to visual fatigue. The fact that no visual fatigue was measured for participants that were less susceptible to visual fatigue may indicate that no visual fatigue was present. It may also indicate that their visual fatigue disappeared more rapidly once the 3D stimulus was gone or that the changes within their visual system were smaller. The fact that no visual fatigue was measured for the nine-view display can be attributed to the lower amount of disparity (within the zone of comfort of one degree of disparity). Nonetheless, the viewers experienced visual discomfort. Although more research is required, it is expected that indeed no actual visual fatigue was induced by this display, but that the high amounts of crosstalk increased the visual discomfort in terms of annoyance.

6 CONCLUSION

We made a first attempt to find the most appropriate indicators, both objective and subjective, for measuring visual fatigue associated with 3D-TV. An initial optometric screening allowed us to find people that were more susceptible to visual fatigue related to 3D-TV. For this group of susceptible people, the only objective indicator that revealed a change within the visual system due to short-term stereoscopic viewing is the fusion amplitude. The subjective indicators double vision and sharpness strongly relate to changes from 2D to 3D, while the subjective indicators pain and irritation strongly relate to the participant's degree of susceptibility to visual fatigue. Hence, we suggest that a combination of fusion range measurement and self-report is appropriate for measuring visual fatigue related to 3D-TV.

7 ACKNOWLEDGEMENTS

We thank Professor Bruce Evans. His proposal for using the algorithm for decompensated heterophoria as a non-arbitrary tool to differentiate in susceptibility to visual fatigue associated with 3D displays provided useful insights in the data. Also, we to thank David Thompson for providing us free usage of Test Chart 2000. And we thank Eelco Pandelaar, Joris Oosterhof and Patrick Werkhoven for their support in the acquisition of data during this study.

8 REFERENCES

[1] Lambooij, M., IJsselsteijn, W. A., Fortuin, M. & Heynderickx, I., "Visual discomfort in stereoscopic displays: a review," Journal of Imaging Science and Technology, submitted.

[2] Peli, E., "The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display," Vision Research, vol. 38, pp. 2053-2066 (1998).

[3] Speranza, F., Tam, W. J., Renaud, R. & Hur, N., "Effect of disparity and motion on visual comfort of stereoscopic images," *Proc. SPIE*, vol. 6055, pp. 60550B (2006).

[4] Yano, S., Emoto, M. & Mitsuhashi, T., "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, pp, 141-150 (2004).

[5] Wopking, M., "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, pp. 101-103 (1995).

[6] Hiruma, N. & Fukuda, T., "Accommodation response to binocular stereoscopic TV images and their viewing conditions," *SMPTE journal*, vol. 102, pp. 1137-1144 (1993).

[7] Emoto, M., Nojiri, Y. & Okano, F., "Changes in fusional vergence limit and its hysteresis after viewing stereoscopic TV," Displays, vol. 25, pp. 67-76 (2004).

[8] Emoto, M., Niida, T. & Okana, F., "Repeated Vergence Adaptation Causes the Decline of Visual Functions in Watching Stereoscopic Television," Journal of Display Technology, vol. 1, pp. 328-340 (2005).

[9] Sheard, C., "The prescription of prism," American Journal of Optometry, vol.11 (**10**): pp. 364-378 (1934).

[10] Evans, B. J. W., "Pickwell's Binocular Vision Anomalies: Investigation and Treatment," Fourth edition, Butterworth-Heinemann, London, UK, (2002).

[11] Borsting, E. J., Mitchell, G. L., Cotter, S. A., Taylor Kulp, M. & The Convergence Insufficiency Treatment Group, "Validity and reliability of the revised convergence insufficiency symptom survey in children aged 9 to 18 years," Optometry Vision Science, vol. 80 (**12**): pp. 832-838 (2003).

[12] Schiffman, R. M., Christianson, M. D., Jacobsen, G., Hirsch, J. D. & Reis, B. L., "Reliability and validity of the ocular surface disease index," Arch. Ophthalmology, vol. 118, pp. 615-621 (2001).

[13] Mangione, C. M., Lee, P. P., Gutierrez, P. R., Spritzer, K., Berry, S., Hays, R. D. & teh National Eye Institute Visual Function Questionnaire Field Test Investigators, "Development of the 25-item National Eye Institute Visual Function Questionnaire," Arch. Ophthalmol, vol. 119, pp. 1050 – 1058 (2001).

[14] Sheedy, J. E. , Hayes, J. & Engle, J., "Is all asthenopia the same?," Optometry and Vision Science, vol. 80, pp. 732-739 (2003).

[15] van Berkel, C. & Clarke, J., "Characterization and optimization of 3D-LCD module design," Proceedings of the SPIE **3012**: 179-187 (1997).

[16] Wilkins, A. J. , Jeanes, R. J., Pumfrey, P. D. and Laskier, M., "Rate of Reading Test: its reliability, and its validity in the assessment of the effects of coloured overlays," Ophthalmic and Physiological Optics, vol 16 (**16)**: pp. 491-497 (1996).

[17] Siegel, M., "Perceptions of crosstalk and the possibility of a zoneless autostereoscopic display," *Proc. SPIE*, vol. 4297, pp. 34-41 (2001).

[18] Kooi, F. L. & Toet, A., "Visual comfort of binocular and 3D displays," *Displays*, vol. 25, pp. 99-108 (2004).

[19] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L., "Multivariate Data Analysis," Upper Saddle River, NJ: Pearson Prentice Hall (2006).