

Subjective Evaluation of Stereoscopic Images: Effects of Camera Parameters and Display Duration

W. A. IJsselsteijn, H. de Ridder, and J. Vliegen

Abstract—In this paper, two experiments are presented that were aimed to investigate the effects of stereoscopic filming parameters and display duration on observers' judgements of naturalness and quality of stereoscopic images. The paper first presents a literature review of temporal factors in stereoscopic vision, with reference to stereoscopic displays. Several studies have indicated an effect of display duration on performance-oriented (criterion based) measures. The experiments reported here were performed to extend the study of display duration from performance to appreciation-oriented measures. In addition, the present study aimed to investigate the effects of manipulating camera separation, convergence distance, and focal length on perceived quality and naturalness.

In the first experiment, using display durations of both 5 and 10 s, 12 observers rated naturalness of depth and quality of depth for stereoscopic still images. The results showed no significant main effect of display duration. A small yet significant shift between naturalness and quality was found for both duration conditions. This result replicated earlier findings, indicating that this is a reliable effect, albeit content-dependent. The second experiment was performed using display durations ranging from 1 to 15 s. The results of this experiment showed a small yet significant effect of display duration. Whereas longer display durations do not have a negative impact on the appreciative scores of optimally reproduced stereoscopic images, observers do give lower judgements to monoscopic images and stereoscopic images with unnatural disparity values as display duration increases. In addition, the results of both experiments provide support for the argument that stereoscopic camera toe-in should be avoided if possible.

Index Terms—Display duration, naturalness, quality, stereoscopic displays, stereoscopic filming.

I. INTRODUCTION

A. Background of Three-Dimensional TV (3DTV) Research

STEREOSCOPIC displays are increasingly being used for both professional and entertainment purposes. A number of studies have shown that 3-D stereoscopic images have a greater

psychological impact, e.g., enhance the viewers' sense of presence [1]–[3], and provide better picture quality than conventional 2-D images [4]–[6]. Although stereoscopic television has received considerable attention in the past, the advent of digital TV transmission reinforces its relevance. Digital transmission enables broadcasters to transmit two synchronized digital channels (one for the left and one for the right eye) in bandwidths smaller than those utilized by one analog TV channel. In addition, stereoscopic display technology has evolved greatly over the past decades [7]–[9], and 3DTV is considered to be the logical next step following HDTV. The current research is motivated by the need to obtain a deeper understanding of the variables that may influence the user's appreciation of stereoscopic images of natural scenes, in the context of stereoscopic television services.

The history of stereoscopy can be traced back to 1838, when Sir Charles Wheatstone presented his classic paper “On some remarkable, and hitherto unobserved, phenomena of binocular vision” to the Royal Society of London. He had developed the “stereoscope,” a device that enabled the viewer to fuse two images of an object or scene into one stereoscopic image, containing a compelling sensation of depth. Sir David Brewster further developed the stereoscope using prisms to magnify and fuse the stereo images. Viewing stereoscopic still images became a popular pastime in Europe and the U.S., and from 1860 to the 1930's stereography flourished. Brewster's lenticular stereoscope became a commercial success, selling 250000 in a short time [10].

The development of motion pictures affected the popularity of the stereoscope, but stereoscopic cinema (early 1900's) and stereoscopic television (1928) were present at the dawn of their monoscopic counterparts [10], [11]. In the 1950's, stereoscopic cinema received much attention from cinematographers. The seminal paper by Spottiswoode, Spottiswoode, and Smith [12] summarized the basic geometrical principles of the three-dimensional film, and introduced concepts such as the “nearness factor,” which facilitated the communication between the film director and stereo technician. Various accounts of the geometry of stereoscopic camera and display systems have been published, both for parallel and toed-in camera configurations [13]–[15]. Although many good stereoscopic movies were produced in the 1950's, stereoscopic cinema got a bad reputation with the public because of the discomfort experienced when viewing misaligned and overdone stereoscopic movies.

More recently, with the growing interest in stereoscopic broadcast television services, a number of laboratories, most notably in Japan, Germany, France, and Canada, have investigated the human factors requirements for a high-quality

Manuscript received March 15, 1999; revised September 30, 1999. This work was completed under workpackage 2 of the TAPESTRIES project, supported under the ACTS initiative of the Commission of European Communities. An earlier version of this paper was presented at the *IS&T/SPIE Conf. Human Vision and Electronic Imaging IV '99*. This paper was recommended by Guest Editor K. N. Ngan.

W. A. IJsselsteijn is with IPO, Center for User-System Interaction, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: W.A.IJsselsteijn@tue.nl).

H. de Ridder is with Delft University of Technology, Industrial Design Engineering, 2628 BX Delft, The Netherlands.

J. Vliegen is with Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.

Publisher Item Identifier S 1051-8215(00)02019-X.

TABLE I
SUMMARY OF THE PERCEPTUAL EFFECTS OF MANIPULATING THE STEREOSCOPIC FILMING PARAMETERS UNDER STUDY
(ADAPTED FROM MILGRAM AND KRÜGER [25])

Parameter	Effect of increase	Effect of decrease
Camera separation (values used in the experiment: 0, 4, 8, 12, 16 and 24 cm)	Increase in disparity values Increase in perceived depth Constant size of objects Constant field of view	Decrease in disparity values Decrease in perceived depth Constant size of objects Constant field of view
Camera focal length (values used: 10 and 20 mm)	Increase in disparity values Increase in perceived depth Increase in size of objects Decrease in field of view	Decrease in disparity values Decrease in perceived depth Decrease in size of objects Increase in field of view
Convergence distance (values used: 1.30 and 2.60 m)	Decrease in disparity values Shift in perceived depth Constant size of objects Constant field of view	Increase in disparity values Shift in perceived depth Constant size of objects Constant field of view

stereoscopic television system. Requirements on viewing conditions (e.g., display size, viewing distance) and picture quality (e.g., tolerances for coding distortions, displacement errors, etc.) have been examined [16]–[19]. In addition, the visual discomfort or eye strain that is sometimes associated with prolonged viewing of stereoscopic images has received attention [4], [20]–[22].

B. Motivation of the Present Study

In the context of stereoscopic televisual services, the experiments presented in this paper were aimed at investigating the effects of stereoscopic filming parameters and display duration on observers' appreciation of stereoscopic images, in particular their judgements of naturalness and quality. The stereoscopic filming parameters under study are summarized in Table I, along with their predicted perceptual effects.

In the context of our research, the main difference between naturalness and quality as subjective evaluation concepts lies in the fact that naturalness refers to what subjects perceive as a truthful representation of reality (i.e., perceptual realism), whereas perceived quality refers to a subjective preference scale. Research in the color domain of image quality has shown that observers are able to differentiate between the two concepts in an experimental situation, and has suggested an interesting relation between image quality and naturalness. De Ridder and colleagues found a small but systematic deviation between image quality and naturalness. This deviation was interpreted to reflect the subjects' preference for more colorful but, at the same time, somewhat unnatural images [23], [24].

Recent preliminary results in the area of stereoscopic image evaluation suggested a similar relation between quality and naturalness for the domain of stereoscopic depth perception. The results suggested that observers preferred (i.e. judged of high quality) a reproduction of stereoscopic depth they also judged to be slightly unnatural. This effect was relatively small for a display duration of 5 s, yet more pronounced for a display duration of 25 s [4].

These results and insights from a subsequent literature review, reported on in Section II, provided motivation for a series of ex-

periments investigating the effects of display duration on quality and naturalness judgements in a more systematic way.

C. Overview of the Paper

First, a literature review relevant to the effects of display duration (or stimulus duration) on stereoscopic vision will be presented. Next, two experiments will be presented. Experiment 1 was aimed to investigate the effect of camera separation, focal length, and convergence distance, as well as display duration on the subjective judgements of quality and naturalness. In Experiment 2, the same independent and dependent variables were employed using, however, a slightly different procedure of stimulus presentation and a wider range of display durations. After discussing the experiments separately, some general conclusions will be drawn.

II. TEMPORAL FACTORS IN STEREOSCOPIC VISION

Temporal factors are known to play a role of considerable importance in the ability of humans to perceive a stereoscopically presented stimulus in depth. A central issue in this area is the minimum image presentation time (stimulus duration or display duration) needed to achieve a depth percept. There have been several investigations focusing on this issue relating it, for instance, to the complexity of the stimulus or the magnitude and direction of the image disparity. The duration of the physical stimulus and the duration of the neural processing are two distinct elements in the literature. These elements are commonly referred to as *stimulus duration* and *stereoscopic processing time*.

Uttal, Davis, and Welke [26] showed that, when using a preconvergence procedure to control fixation disparity at stimulus onset, a compelling stereoscopic experience could be elicited for different random-dot stereoscopic forms with very brief (<1 ms) stimulus durations. They conclude that higher temporal threshold estimates found elsewhere in the literature are more likely a measure of the time it takes to converge the eyes and establish correspondence between the two images, rather than the minimum time a stereoscopic stimulus must be presented before a depth percept occurs.

However, it takes time to process the information in a stereogram after such a brief exposure. Julesz [27], [28] varied the interstimulus interval between the brief presentation of a simple unambiguous random-dot stereogram and the presentation of an ambiguous random-dot stereogram. While the unambiguous stereogram contained either crossed (i.e., depth in front of fixation) or uncrossed (i.e., depth behind fixation) disparity information, the ambiguous stereogram contained disparities in both directions. With presentation times over 50 ms, it was found that the second (ambiguous) stimulus was consistently perceived at the same depth as the first (unambiguous) stimulus. When the interstimulus interval was less than 50 ms, the unambiguous random-dot stereogram did not bias the interpretation of the ambiguous stereogram.

Uttal, Fitzgerald, and Eskin [29] have obtained similar results as Julesz [27], [28] by using poststimulus masking with varying intervals between the stimulus and the mask. Performance was degraded when the blanking field followed the test stimulus by less than about 50 ms, suggesting that 50 ms would be the time necessary to establish a perception of depth.

In general, psychophysical research in this area has been performed using well-trained observers that readily see depth in stereograms. However, it has been shown that large individual differences exist between subjects, and that many people have difficulty seeing depth in random-dot stereograms and may require many seconds or even minutes of exposure to the stimulus before they see depth. Relatedly, the percentage of the population judged to be stereoanomalous (i.e., have a deficit in stereopsis) is dependent on the testing method used and more specifically on the stimulus duration that is used for a test [30], [31].

Tam and Stelmach [31] systematically varied display duration until viewers were able to perform a depth discrimination task with 75% accuracy. Their results showed that the proportion of observers unable to perform the depth discrimination task declined with increasing display duration. At the shortest display duration (20 ms) approximately half of the observers ($N = 100$) were unable to perform the stereoscopic depth discrimination task at criterion level. However, at the maximum display duration of 1000 ms, failure rate dropped to about 5%. These results clearly indicate that the prevalence of stereoanomaly critically depends on the display duration with which the stereoscopic stimuli are shown. One important implication for stereoscopic televisual services of these findings is that about 95% of the population is expected to experience an enhanced sense of depth when viewing a stereoscopic program, since most scene durations typically exceed one second between cuts or fades, which is a sufficiently long duration for most people.

An interesting issue is the impact temporal factors may have on the *judgement* of depth in stereoscopic media. In this context, it is useful to distinguish between a *performance-oriented* approach and an *appreciation-oriented* approach [4], [32].

Within a *performance-oriented approach*, the accuracy of depth discrimination is the most fundamental concern, especially when stereoscopic displays are being used for guiding high-precision manipulations, such as stereoscopic endoscopy or remote guidance and inspection in hazardous industrial settings. In a direct investigation of several factors that may affect the veridicality of depth perception using stereoscopic

displays, Patterson, Moe & Hewitt [33] used a brief (160 ms) and a long (unlimited) display duration to test whether depth is perceived as would be expected on the basis of stereoscopic geometry. For the long display duration, the depth judgements followed the geometrically predicted depth quite closely, especially for depth in the crossed disparity direction. For the brief display duration, depth judgements were less accurate, deviating especially in conditions with uncrossed disparity, large half-image separation, and/or a long viewing distance.

Patterson, Bowd, Becker, Monaghan, Shorter, and Gilbert [34] expanded on this work, investigating disparity scaling via a task involving depth discrimination. They used crossed and uncrossed directions of disparity combined with four display durations (67, 167, 417, and 5000 ms). Their results showed that the percentage correct discrimination averaged across 100 observers (25/display duration condition) increases with an increase in stimulus duration, with percentage correct discrimination higher for the crossed disparity direction. Thus, when display duration was long, most individuals discriminated depth correctly in both directions, but specifically in the crossed direction. At brief display durations, the percentage correct discrimination decreased markedly, especially for the uncrossed direction, where at 67 ms the percentage correct was less than 50%. These results are in line with the earlier results of Patterson *et al.* [33], [35] and the results of Tam and Stelmach [31]. They indicate a build-up of the depth percept over time, varying across individuals.

The effect that display duration may have on the subjective *appreciative* judgements of stereoscopic pictures depicting natural scenes has received very little attention to date. The current paper aims to extend the study of exposure duration from performance (criterion-based) measures to appreciation measures, using realistic stereoscopic images. In theory, one would expect display duration to be relevant for subjective appreciation in three ways. First, a minimal duration is necessary to allow proper vergence responses of the eyes (typically 100 ms). Secondly, the inspection of a stereoscopic scene, or build-up of the stereoscopic depth structure, would presumably take more time than inspecting a monoscopic image, since more information is present in the stereoscopic case and, in general, stereovision has a long temporal integration period, relative to other visual functions [36], [37]. Thirdly, one would expect an adverse effect of display duration on appreciative judgements when image parallaxes are excessive. These are generally found to be visually disturbing, potentially causing eye strain and headache [4], [20]–[22].

In sum, the literature reviewed in the current section¹ clearly demonstrates the impact that temporal factors have on stereoscopic vision. It has been demonstrated that the stimulus duration necessary for disparity detection can be very brief (<1 ms), provided the right kind of stimulus is used and the experimental conditions are well controlled. The minimal processing time of a stereoscopic stimulus is substantially longer (around 50 ms). Additionally, the time necessary to build up a depth percept and to be able to discriminate, for instance, between crossed and

¹A more comprehensive literature review relating to this topic is provided elsewhere [38]. This document is available on request from the corresponding author.

uncrossed depth, is variably distributed across the population. In the remainder of the current paper, two experiments will be presented that were designed to investigate the effect of display duration on the subjective appreciation of stereoscopic images.

III. EXPERIMENT 1

A. Introduction

This experiment was performed for two reasons mainly. First, we wanted to investigate if and how quality and naturalness judgements would be affected by manipulating stereoscopic filming parameters and display duration, thus extending the study of stereoscopic display duration from performance to appreciation oriented measures. Secondly, we wanted to check the reliability of the results reported elsewhere [4], in particular the small but systematic quality-naturalness shift that was observed.

B. Method

1) *Observers*: Twelve observers (6 male, 6 female, average age 25, range 20–36) either volunteered or were paid to participate in the experiment. All observers were naïve with respect to the experimental hypothesis being tested. All had normal or corrected to normal visual acuity (a visus of ≥ 1 as tested on the Landolt C test) and a stereoacuity of 30 sec-arc or better (as tested on the RANDOT© random dot stereotest, Stereo Optical Co., Inc., Illinois).

2) *Apparatus*: The stereoscopic image pairs were displayed on an AEA Technology 20" stereoscopic display consisting of two BARCO CPM 2053FS color monitors (50-Hz PAL), with polarized filters in front of each (see Fig. 1). Observers viewed the display wearing polarized spectacles to separate right and left eye view. The video input to the display was provided by a SUN ISP system running custom software to control display duration and synchronize the output of the 2 codecs transferring the images.

3) *Stimuli*: The image material consisted of stereoscopic image pairs varying in camera separation, focal length and convergence distance. Two levels of focal length were employed, 10 and 20 mm, as well as two levels of convergence distance, 1.30 and 2.60 m. The following combinations between focal length and convergence distance were used: $F : 10 \text{ mm } C : 1.30 \text{ m}$, $F : 20 \text{ mm } C : 1.30 \text{ m}$, and $F : 20 \text{ mm } C : 2.60 \text{ m}$. Within each of these three combinations, six levels of camera separation were used, based on the distance between the right and left camera: 0, 4, 8, 12, 16, and 24 cm. This resulted in a total of 18 unique stimuli per scene, varying in stereoscopic filming parameters. The images used depicted two different scenes. The scene, called "Playmobiles," consisted of a colorful toy landscape with mountains and numerous Playmobiles. The other scene, called "Bureau," consisted of a tailor's dummy sitting behind a desk on which some office equipment is located. The stereoscopic images were kindly provided to us by CCETT, France.

4) *Procedure*: Observers were seated at a viewing distance of 0.80 m (approximately 2.5 picture heights) from the stereoscopic display. They were given written instructions detailing the task they had to perform, and the attribute they were asked

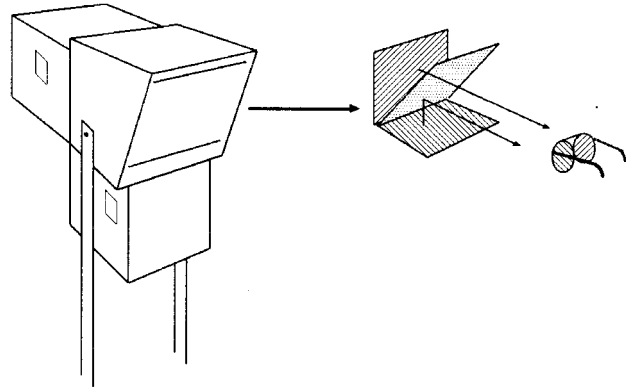


Fig. 1. Schematic drawing of the AEA dual-monitor stereoscopic display used in the reported experiments. It presents the right and left eye images at the same time (i.e., time parallel), using polarization to separate both views.

to rate. These instructions were then reiterated by the experimenter as to ensure the observer understood the task at hand.

The stereoscopic images were presented on the 20" stereodisplay, placed in a dimly lit laboratory. The 36 images (18 stereoscopic parameter variations * 2 different scenes) were presented in random order and subsequently in reversed randomized order to compensate for potential effects of presentation order (such as adaptation effects or fatigue). The 5- and 10-s display duration conditions were presented in separate counterbalanced sessions, separated by one day.

For each of the durations, observers were asked to rate *quality* of depth and *naturalness* of depth in subsequent counterbalanced sessions using a scale from one to ten, where one represents the lowest level and ten represents the highest level of the scaled attribute (magnitude estimation). They could enter their attribute rating on the numerical part of a keyboard placed in front of them. Before the experiment started the observers were shown a small set of practice stimuli displaying the range of camera parameter variations over which they had to provide a rating. During the experiment, a grey adaptation field was shown between two successive stimuli, to allow the eyes to return to a resting state. The adaptation time was determined by the time that observers took to input their response, but with a minimum of 5 s.

C. Results

Fig. 2 shows the quality and naturalness ratings for each of the stereoscopic filming parameters. A separate graph is presented for each combination of focal length and convergence distance used. The top and bottom panel show the results for the 5- and 10-s display duration conditions respectively. These graphs show the average over 12 observers, two scenes, and two presentation orders (i.e., the original and the reversed randomized sequence), with the error bars reflecting the standard error of the mean of each data point. Camera-base distance is plotted on the x -axis in centimeters, the attribute scores are plotted on the y -axis.

Fig. 2 reveals a clear effect of camera-base distance and of the focal lengths and convergence distances used. A repeated measures analysis of variance was performed which showed a significant main effect of camera base distance ($F(2, 22) =$

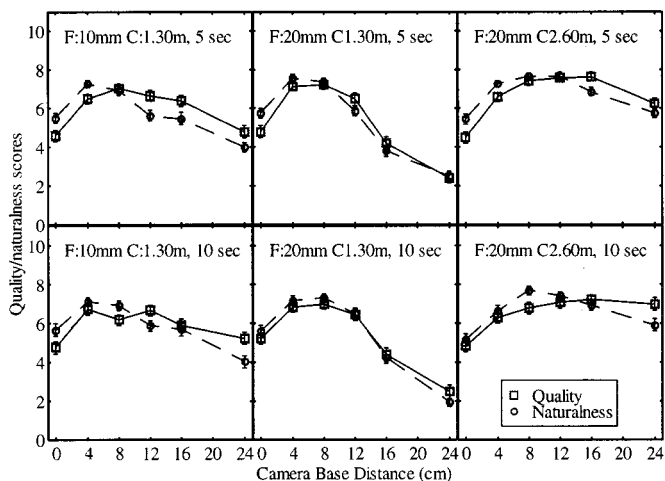


Fig. 2. Results of Experiment 1—quality and naturalness scores plotted against camera base distance (in cm). The graphs in the top panel show the results of the 5-s display duration condition, the graphs in the bottom panel show the results for the 10-s condition.

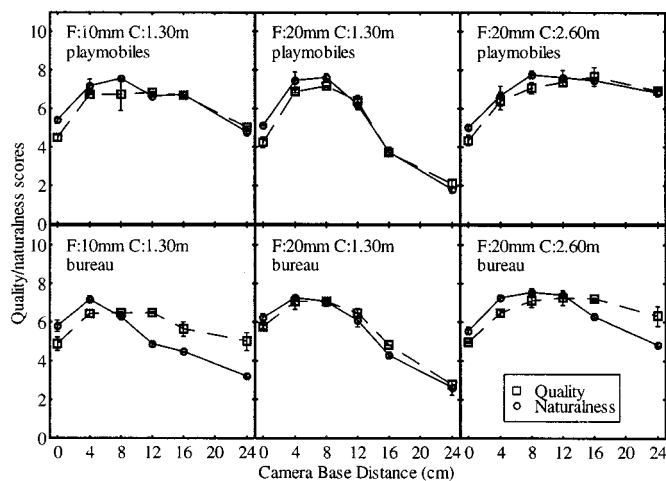


Fig. 3. Results of Experiment 1—quality and naturalness scores plotted against camera base distance (in cm). The graphs in the top panel show the results for the scene “Playmobiles,” the graphs in the bottom panel show the results for the scene “Bureau.”

32.43, $p < 0.01$) and of camera condition, i.e., the three combinations of focal length and convergence distance, ($F(5, 55) = 19.37, p < 0.01$). Not surprisingly, a significant interaction ($F(10, 110) = 21.52, p < 0.01$) between camera separation and camera condition was also found. No significant main effect was found of display duration. However, a significant interaction was found between display duration and scene ($F(1, 11) = 6.03, p < 0.05$). A pairwise comparison per scene revealed that the effect of display duration was significant ($p < 0.05$) for the “Playmobiles” scene, but not for the “Bureau” scene. In particular, for the “Playmobiles” scene, people tended to give somewhat lower ratings at longer durations. This suggests a duration effect that is content dependent.

Fig. 2 also illustrates that although quality and naturalness judgements are highly correlated (Pearson's r of 0.89), a small but systematic shift can be seen between the quality and naturalness scores for the camera conditions $F : 10$ mm and $C : 1.30$ m and $F : 20$ mm, $C : 2.60$ m. For low values of camera separation (4 and 8 cm), naturalness is rated slightly higher than quality, whereas with increasing camera separation, quality is rated higher. This effect is reflected in the significant two-way interaction between task (i.e., rating of quality or naturalness) and the camera separation ($F(5, 55) = 6.76, p < 0.01$), as well as the significant three-way interaction between task, camera separation, and camera condition ($F(10, 110) = 2.35, p < 0.02$).

However, this effect is only clearly found for one scene (i.e., “Bureau”), as is illustrated in Fig. 3. The relevant interactions are approaching significance: a three-way interaction between scene, task, and camera separation ($F(5, 55) = 2.17, p = 0.07$) and a four-way interaction between scene, task, camera separation, and camera condition ($F(10, 110) = 1.87, p = 0.056$).

D. Discussion

Both the quality and naturalness judgements (see Figs. 2 and 3) clearly reflect the effect of changes in disparity values. As summarized in Table I, an increase in camera separation

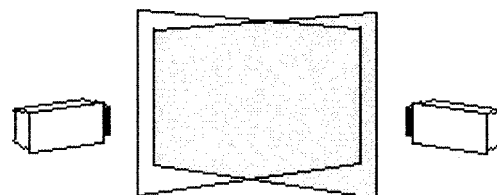


Fig. 4. A schematic representation of the trapezoid projections using a toed-in camera configuration.

and in focal length of the lens result in an increase in disparity values, whereas with an increase in convergence distance disparity values decrease. Both quality and naturalness of depth increase with the transition from monoscopic (0-cm camera base distance) to stereoscopic mode of presentation, suggesting observers prefer stereoscopic presentation of images, provided the disparities are kept within natural bounds, as evidenced by the drop in quality and naturalness scores at high camera separation values. This drop is especially noticeable for the condition with the long focal length and short convergence distance ($F : 20, C : 1.30$), in which the disparity values are most extreme.

One might expect there to be little difference between the $F : 10, C : 1.30$ condition and the $F : 20, C : 2.60$ condition, since the increase in horizontal disparity values caused by a longer focal length is compensated by the decrease in disparity values caused by the longer convergence distance. However, the quality and naturalness scores are higher for camera separations larger than 8 cm. when comparing the $F : 20, C : 2.60$ condition to the $F : 10, C : 1.30$ condition. This is likely due to the fact that with a toed-in or converging camera configuration keystone distortion may occur. Keystone distortion is an image distortion caused by vertical parallax in the stereoscopic image due to the fact that the imaging sensors of the cameras are directed toward slightly different planes [14], resulting in two slightly trapezoid projections (when projected parallel) oriented in opposite directions for each of the cameras (see Fig. 4).

The amount of vertical disparity is greatest in the corners of the image and increases with increased camera separation, decreased convergence distance and decreased focal length [14]. Perceptually, this may result in a depth plane curvature, which may have had a negative impact on the observer's appreciative judgements. Our results thus provide a subjective basis for the argument that camera toe-in should be avoided in stereoscopic filming. This argument is also in line with the results presented by Yamanoue [39], who showed that a toed-in camera configuration may lead to a size distortion known as the puppet theater effect, whereas parallel camera configurations do not produce such size distortions.

The quality-naturalness shift reported earlier [4] was replicated in the current experiment and reflected in the significant two and three-way interactions reported. At closer inspection of the data per scene, we find that the overall shift is mainly attributable to one of the two scenes ('bureau'). This would indicate that although the quality-naturalness shift is a reliable effect, it also seems to be content-dependent. Speculatively, since the monocular depth cues (e.g., texture, occlusion, contour, perspective, shading) that are available in each of the two scenes differ considerably, this may have had an effect on the relative 'added value' or weighting of the stereoscopic depth cue in the total percept and the subsequent appreciation of the scene. This explanation would be in line with current depth cue integration theory suggesting that our depth percept is a weighted sum of the different available depth cues [40], [41]. However, to properly investigate such a potential explanation in the context of realistic stereoscopic images, an experiment would be required in which a substantial number of content variations are introduced in a systematic and well-controlled manner. Such an experiment was beyond the scope of the present paper.

Importantly, the results did not show a significant main effect of display duration, indicating that averaged over the two different scenes observers did not differentiate in their appreciative judgements of the stereoscopic scenes between the 5- and 10-s display duration conditions. The reason for not finding a significant main effect of display duration might be that the range of display durations was either too small or too one-sided; both 5 and 10 s can be regarded as fairly long display durations. To investigate this possibility, a second experiment using a wider range of display durations was performed.

IV. EXPERIMENT 2

A. Introduction

The rationale for carrying out Experiment 2 was to further investigate the potential effect of display duration on appreciative judgements of stereoscopic images that varied in stereoscopic filming parameters using a wider range of display durations than the two conditions used in Experiment 1. In this experiment, display durations of 1, 3, 5, 10, and 15 s were employed.

B. Method

1) *Observers:* Twelve observers (10 male, 2 female, average age 26.3, range 21–39) either volunteered or were paid to participate in the experiment. All observers were naïve with respect to the experimental hypothesis being tested. All had normal or

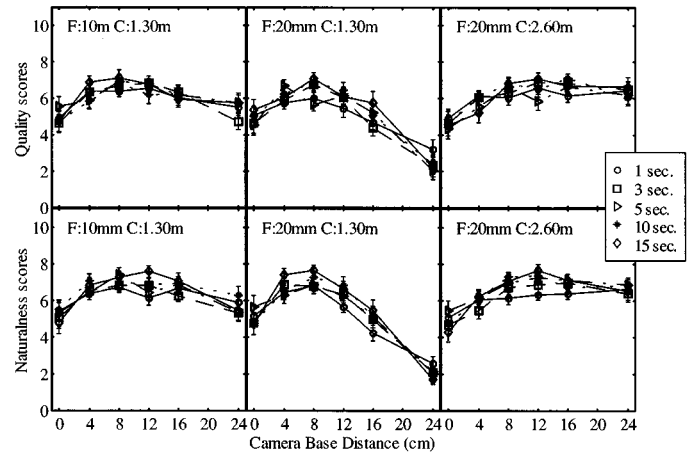


Fig. 5. Results of Experiment 2. Quality (top) and naturalness (bottom) scores as a function of camera separation.

corrected to normal visual acuity (a visus of ≥ 1 as tested on the Landolt C test) and a stereoacuity of 30 sec-arc or better (as tested on the RANDOT© random dot stereotest, Stereo Optical Co., Inc., Illinois).

2) *Apparatus:* The apparatus was identical to the equipment used in Experiment 1.

3) *Stimuli:* The image material used as stimulus set was identical to the stimuli used in Experiment 1 in terms of the stereoscopic parameters. However, only one scene ("Playmobiles") was used, mainly to limit the duration of the total experiment.

4) *Procedure:* The procedure was largely identical to the procedure in Experiment 1, however with a few important differences. First, the range of display durations used was extended to 1, 3, 5, 10, and 15 seconds of display duration. Secondly, stimuli presented at these different durations were rated within one session, instead of in separate sessions as was the case in the first experiment. The stimuli were fully randomized over both stereoscopic parameters and all display durations and, as with Experiment 1, this random order was subsequently presented in reverse order as well. Observers were asked to rate *quality* of depth and *naturalness* of depth in different counterbalanced sessions separated by at least one day.

C. Results

Fig. 5 shows the averaged quality rating (top panel) and naturalness ratings (bottom panel) for each of the stereoscopic filming parameters. A separate graph is presented for each combination of focal length and convergence distance. Each line represents the results for a different display duration. These graphs show the average over 12 observers and 2 presentation orders (i.e., the original and reversed randomized sequence), with the error bars reflecting the standard error of the mean of each data point. Camera-base distance is plotted on the x -axis in centimeters, the attribute scores are plotted on the y -axis.

A repeated measures analysis of variance was performed with the quality/naturalness scores as a dependent variable and stimulus duration, task (i.e. rating of quality or naturalness), camera condition (i.e., the combination of focal length and convergence distance) and camera separation as factors.

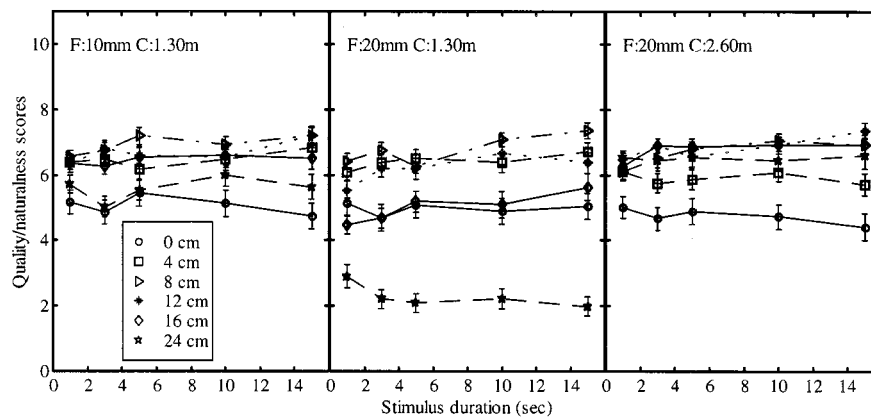


Fig. 6. Pooled quality/naturalness scores as a function of stimulus duration for each camera condition and camera separation.

Again, and not surprisingly, camera condition and separation had a significant main effect ($F(2, 22) = 13.68, p < 0.01$ and $F(5, 55) = 4.78, p < 0.01$, respectively), as well as a significant interaction ($F(10, 110) = 16.38, p < 0.01$). The results showed no clear quality-naturalness shift, nor was this suggested by an interaction between task, camera separation, and/or camera condition.

We found a significant main effect of stimulus duration ($F(4, 44) = 2.65, p < 0.05$), as well as a significant interaction between stimulus duration and camera separation ($F(20, 220), p < 0.01$). The trends of these effects are illustrated in Fig. 6, which shows the pooled quality and naturalness scores plotted against stimulus duration, with each line representing a different camera separation.

D. Discussion

As with Experiment 1 and earlier experiments, changes in disparity values highly affect the quality and naturalness judgements, and determine to a large extent the shape of the curve. Again, observers preferred stereoscopic images with natural disparities over monoscopic images and images projected with extreme disparity values. This proves to be a reliable and robust finding. The small but systematic quality-naturalness shift that we found in our earlier experiments wasn't replicated in Experiment 2. However, the results of Experiment 1 showed that this shift may be content-dependent, and indeed was not clearly present in the results for the scene that we used for Experiment 2 (i.e., “playmobiles”).

The effect of display duration was found to be small yet significant. It is best illustrated by Fig. 6, which shows a slight divergence of the results for the different camera separations as display duration increases. In particular, the averaged scores for the 0- and 24-cm camera separation tend to decrease with increasing display duration, whereas a slight increase can be observed for some of the other camera separations. This divergence is most salient for the condition with a long focal length and a short convergence distance. So, whereas longer display durations do not have a negative impact on the appreciative scores of optimally reproduced stereoscopic images, observers

do give lower judgements to monoscopic images and stereoscopic images with unnatural disparity values as display duration increases.

Although the effect of exposure duration was small, it is possible that a more pronounced duration effect would be observed for durations under 1 s. However, for the present purpose of stereoscopic television, most scenes (between cuts or fades) will have a length of more than 1 s. Our results have shown that observers are equally able to appreciate the depth in stereoscopic scenes presented for only 1 s, as they are able to judge scenes of a longer duration.

V. GENERAL CONCLUSIONS

Several studies have indicated a critical effect of display duration on performance-oriented (criterion based) measures. The experiments reported in this paper were performed to extend the study of display duration from performance to appreciation-oriented measures. In addition, the present study aimed to investigate the effects of manipulating camera separation, convergence distance and focal length on perceived quality and naturalness, and the relation between these two dependent variables.

The results of the presented experiments support a number of conclusions. First of all, the effect that display duration has on the subjective evaluation of stereoscopic images is relatively small. Secondly, the quality-naturalness shift that was found in an earlier experiment [4] was replicated in Experiment 1, although mainly in connection with one particular scene. Thus, this effect seems to be reliable yet content-dependent. This indicates, however, that optimal naturalness need not necessarily coincide with optimal quality, a result that is in line with the computational theory of image quality, put forward by Janssen and Blommaert [42], that suggests that for image quality judgements both naturalness and usefulness are weighted in. Thirdly, the results of both experiments provide support for the argument that stereoscopic camera toe-in should be avoided if possible. Finally, and not surprisingly, our results seem to indicate a preference

of observers for stereoscopically presented images, provided the disparity values are kept within natural bounds.

ACKNOWLEDGMENT

The authors gratefully acknowledge their partners in the TAPESTRIES project, in particular CCETT in France, for providing the stereoscopic stimuli. In addition, the authors acknowledge with appreciation Jonathan Freeman (University of London, Goldsmith College, U.K.) and Steve Avons (Department of Psychology, University of Essex, U.K.). They also wish to thank Martin Boschman, who wrote the custom software needed to control the display duration of the stereoscopic stimuli. Dirk Suykens, a visiting student from Brussels Free University, Belgium, was kind enough to administer the first experiment.

REFERENCES

- [1] C. Hendrix and W. Barfield, "Presence within virtual environments as a function of visual display parameters," *Presence: Teleoperators and Virtual Environ.*, vol. 5, pp. 274–289, 1996.
- [2] W. A. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," *Displays*, vol. 18, pp. 207–214, 1998.
- [3] J. Freeman, S. E. Avons, D. Pearson, and W. A. IJsselsteijn, "Effects of sensory information and prior experience on direct subjective ratings of presence," *Presence: Teleoperators and Virtual Environ.*, vol. 8, pp. 1–13, 1999.
- [4] W. A. IJsselsteijn, H. de Ridder, and R. Hamberg, "Perceptual factors in stereoscopic displays. The effect of stereoscopic filming parameters on perceived quality and reported eye-strain," *Proc. SPIE*, vol. 3299, pp. 282–291, 1998.
- [5] S. Yano and I. Yuyama, "Stereoscopic HDTV: Experimental system and psychological effects," *SMPTE J.*, vol. 100, pp. 14–18, 1991.
- [6] A. Berthold, "The Influence of Blur on the Perceived Quality and Sensation of Depth of 2D and Stereo Images," ATR Human Information Processing Research Laboratories, Kyoto, Japan, Tech. Rep. H-232, 1997.
- [7] T. Okoshi, "Three-dimensional displays," *Proc. IEEE*, vol. 68, pp. 548–564, 1980.
- [8] S. Pastoor and M. Wöpking, "3-D displays: A review of current technologies," *Displays*, pp. 100–110, 1997.
- [9] I. Sexton and P. Surman, "Stereoscopic and autostereoscopic display systems. An in-depth review of past, present, and future technologies," *IEEE Signal Processing Mag.*, pp. 85–99, May 1999.
- [10] R. Zone, "The deep image: 3D in art and science," *Proc. SPIE*, vol. 2653, pp. 4–8, 1996.
- [11] Radio News (1928, Nov.). [Online]. Available: [Online]. Available HTTP: <http://www.antiqueradios.com/stereotv.shtml>
- [12] R. Spottiswoode, N. L. Spottiswoode, and C. Smith, "Basic principles of the three-dimensional film," *SMPTE J.*, vol. 59, pp. 249–286, 1952.
- [13] J. T. Rule, "The geometry of stereoscopic pictures," *J. Opt. Soc. Amer.*, vol. 31, pp. 325–334, 1941.
- [14] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE*, vol. 1915, pp. 36–48, 1993.
- [15] R. Franich, "Disparity Estimation in Stereoscopic Digital Images," Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, 1996.
- [16] S. Pastoor, "3D-television: A survey of recent results on subjective requirements," *Signal Processing: Image Commun.*, vol. 4, pp. 21–32, 1991.
- [17] T. Motoki, H. Isono, and I. Yuyama, "Present status of three-dimensional television research," *Proc. IEEE*, vol. 83, pp. 1009–1021, 1995.
- [18] L. B. Stelmach, W. J. Tam, and D. Meegan, "Perceptual basis of stereoscopic video," *Proc. SPIE*, vol. 3639, pp. 260–265, 1999.
- [19] S. Yano, "Perception of sensation of reality based on human information processing," in *3D Image Conf. '99*, 1999, pp. 199–206.
- [20] H. Ohzu and K. Habara, "Behind the scenes of virtual reality: Vision and motion," *Proc. IEEE*, vol. 84, pp. 782–798, 1996.
- [21] N. Hiruma, "Accommodation response to binocular stereoscopic TV images," in *Human Factors in Organizational Design and Management—III*, K. Noro and O. Brown, Eds., 1990, pp. 233–236.
- [22] T. Inoue and H. Ohzu, "Accommodation and convergence when looking at binocular 3D images," in *Human Factors in Organizational Design and Management—III*, K. Noro and O. Brown, Eds., 1990, pp. 249–252.
- [23] H. de Ridder, F. J. J. Blommaert, and E. A. Fedorovskaya, "Naturalness and image quality: Chroma and hue variations in color images of natural scenes," *Proc. SPIE*, vol. 2411, pp. 51–61, 1995.
- [24] H. de Ridder, "Naturalness and image quality: Saturation and lightness variations in color images of natural scenes," *J. Imaging Sci. Technol.*, vol. 40, pp. 487–493, 1996.
- [25] P. Milgram and M. Krüger, "Adaptation effects in stereo due to on-line changes in camera configuration," *Proc. SPIE*, vol. 1669, pp. 122–134, 1992.
- [26] W. R. Uttal, N. S. Davis, and C. Welke, "Stereoscopic perception with brief exposures," *Percept. and Psychophys.*, vol. 56, pp. 599–604, 1994.
- [27] B. Julesz, "Binocular depth perception with familiarity cues," *Science*, vol. 145, pp. 356–362, 1964.
- [28] B. Julesz, "Texture and visual perception," *Sci. Amer.*, vol. 212, pp. 38–48, 1965.
- [29] W. R. Uttal, J. Fitzgerald, and T. E. Eskin, "Parameters of tachoscopic stereopsis," *Vis. Res.*, vol. 15, pp. 705–712, 1975.
- [30] R. Patterson and R. Fox, "The effect of testing method on stereoanomaly," *Vis. Res.*, vol. 24, pp. 403–408, 1984.
- [31] W. J. Tam and L. B. Stelmach, "Display duration and stereoscopic depth discrimination," *Can. J. Experimental Psych.*, vol. 52, pp. 56–61, 1998.
- [32] J. A. J. Roufs, "Perceptual image quality: Concept and measurement," *Philips J. Res.*, vol. 47, pp. 35–62, 1992.
- [33] R. Patterson, L. Moe, and T. Hewitt, "Factors that affect depth perception in stereoscopic displays," *Human Factors*, vol. 34, pp. 655–667, 1992.
- [34] R. Patterson, C. Bowd, S. Becker, M. Monaghan, S. Shorter, and J. Gilbert, "Stereoscopic depth discrimination in the crossed and uncrossed directions with brief- vs. extended-stimulus exposure," *SID 96 Dig.*, pp. 973–975, 1996.
- [35] R. Patterson, R. Cayko, L. Short, R. Flanagan, L. Moe, E. Taylor, and P. Day, "Temporal integration differences between crossed and uncrossed stereoscopic mechanisms," *Percept. and Psychophys.*, vol. 57, pp. 891–897, 1995.
- [36] K. I. Beverly and D. Regan, "Temporal integration of disparity information in stereoscopic perception," *Experimental Brain Res.*, vol. 19, pp. 228–232, 1974.
- [37] R. van Ee and C. J. Erkelens, "Temporal aspects of binocular slant perception," *Vis. Res.*, vol. 36, pp. 43–51, 1996.
- [38] W. A. IJsselsteijn, "Temporal factors in stereoscopic vision: A selective literature review," in *Display and Image Parameters Affecting Visual Realism and Presence in New Display Technologies*, W. A. IJsselsteijn, J. Freeman, H. de Ridder, and S. E. Avons, Eds. Eindhoven, The Netherlands: IPO, pp. 81–97.
- [39] H. Yamanoue, "The relation between size distortion and shooting conditions for stereoscopic images," *SMPTE J.*, vol. 106, pp. 225–232, 1997.
- [40] J. P. Frisby, D. Buckley, and J. Freeman, "Stereo and texture cue integration in the perception of planar and curved large real surfaces," in *Attention and Perform.*, T. Inui and J. L. McClelland, Eds., 1996, vol. XVI, pp. 71–91.
- [41] E. B. Johnston, B. G. Cumming, and M. S. Landy, "Integration of stereopsis and motion shape cues," *Vis. Res.*, vol. 34, pp. 2259–2275, 1994.
- [42] T. J. W. M. Janssen and F. J. J. Blommaert, "Image quality semantics," *J. Imaging Sci. Technol.*, vol. 41, pp. 555–560, 1997.



Wijnand A. IJsselsteijn received the Masters degree in psychology from the University of Utrecht, the Netherlands. He is currently working toward the Ph.D. degree in the area of new media evaluation.

After a short career in information technology, he joined Institute for Perception Research, Eindhoven, the Netherlands, in 1996, where his work focused on the European Commission (EC)-funded ACTS project TAPESTRIES. During this project, he worked on the psychological evaluation of new display and communication media. His research interests include stereoscopic displays, multisensory perception, and presence—its structure, determinants, measurement methodologies, and associated technologies.



Huib de Ridder received the M.Sc. degree in psychology from the University of Amsterdam, Holland, and the Ph.D. degree from Eindhoven University of Technology, the Netherlands.

Since 1992, he has been affiliated with the Vision Group of the Institute for Perception Research, Eindhoven, The Netherlands, where his research focused on both fundamental and applied psychophysics. During 1987–1992, he conducted research on the fundamentals of perceptual image quality metrics, supported by a personal fellowship from the Royal Netherlands Academy of Arts and Sciences. In November 1998, he was appointed Associate Professor of Informational Ergonomics, Department of Industrial Design, Delft University of Technology, The Netherlands. His current research interests include image quality metrics, stereoscopic displays, augmented reality, user-product interaction, and interaction with embedded intelligence.



Joyce Vliegen was a student of phonetics at the University of Utrecht, the Netherlands. Her Masters thesis involved psychoacoustical research on auditory streaming at the Institute for Perception Research (IPO), Eindhoven, The Netherlands, and later at the University of Cambridge, U.K. She returned to IPO to work on the ACTS project TAPESTRIES, and in January 1999, began a Ph.D. project at the Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

Her research interests include auditory streaming, auditory sentence perception and the role of prosody, data analysis, and audio-visual interactions.