Viewing experience and naturalness of 3D images

Pieter J.H. Seuntiëns[a], Ingrid E.J. Heynderickx[b], Wijnand A. IJsselsteijn[a], Paul M.J. van den Avoort[a], Jelle Berentsen[a], Iwan J. Dalm[a], Marc T.M. Lambooij[a], Willem Oosting[a]

[a]University of Technology, Den Dolech 2, 5600 MB, Eindhoven, The Netherlands
[b]Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands

ABSTRACT

The term 'image quality' is often used to measure the performance of an imaging system. Recent research showed however that image quality may not be the most appropriate term to capture the evaluative processes associated with experiencing 3D images. The added value of depth in 3D images is clearly recognized when viewers judge image quality of unimpaired 3D images against their 2D counterparts. However, when viewers are asked to rate image quality of impaired 2D and 3D images, the image quality results for both 2D and 3D images are mainly determined by the introduced artefacts, and the addition of depth in the 3D images is hardly accounted for. In this experiment we applied and tested the more general evaluative concepts of 'naturalness' and 'viewing experience'. It was hypothesized that these concepts would better reflect the added value of depth in 3D images. Four scenes were used varying in dimension (2D and 3D) and noise level (6 levels of white gaussian noise). Results showed that both viewing experience and naturalness were rated higher in 3D than in 2D when the same noise level was visible. Thus, the added value of depth is clearly demonstrated when the concepts of viewing experience and naturalness are being evaluated. The added value of 3D over 2D, expressed in noise level, was 2 dB for viewing experience and 4 dB for naturalness, indicating that naturalness appears the more sensitive evaluative concept for demonstrating the psychological impact of 3D displays.

Keywords: 3D image, viewing experience, naturalness, multi-view display, gaussian noise, evaluation concept

1. INTRODUCTION

Since the introduction of the television, much has been done to improve viewing experience. A logical next step is the introduction of 3D content. Proponents of 3D-TV have argued that 3D-TV will bring the viewer a whole new experience, a fundamental change in the character of the image, not just an enhancement of quality.[1, 2]

One of the major cues of depth is the binocular cue. Because the human eyes are located at a slightly different position in the human head, they both receive a different image. The brain fuses these different images to one image and extracts depth information from the difference between the two.[3] Different technologies exist for simulating the binocular depth cue from 2D images. Most common is the system where two images are displayed separately (side-by-side) and optical devices like mirrors or lenses direct the images to the appropriate eye. Another technique displays left and right images on the screen and red-green color filters or polarized filters are used to determine which image should be received by which eye. These systems, where the viewer needs to wear an optical device are known as aided viewing. Relatively new systems can be viewed without the use of any optical devices. These systems are called free viewing- or auto-stereoscopic systems.

Comparisons between television sets are done quite regularly on perceptual and/or technical aspects to determine where to put future investments. The performance of a 3D television system is often evaluated using 2D image quality models.[4] Perceived image quality is considered to be a multidimensional attribute. Earlier research in this area defined some dominant perceptual factors affecting 2D image quality, for instance, blur, blockiness, or noise. Psychophysical scaling experiments are used to quantify the strengths of these artefacts. People use perceptual rules to combine the measured strengths into a prediction of the overall image quality. We believe that 2D image quality models are not adequate to measure 3D image quality because typical stereoscopic

Further author information: (Send correspondence to Pieter J.H. Seuntiëns)
Pieter J.H. Seuntiëns: E-mail: P.J.H.Seuntiens@tue.nl, Telephone: +31 (0)40 247 3531

distortions and the depth reproduction are not corporated. Typical stereoscopic distortions only appear in 3D material such as keystone distortion, puppet theater effect, cardboard effect, shear distortion, picket-fence effect, image flipping and the most important one crosstalk. The positive contribution of depth to perceived image quality was demonstrated for uncompressed 3D images compared to 2D images.[5] However, other research has shown that perceived image quality of MPEG and JPEG coded 3D material was mainly determined by the introduced artefacts and not by the depth reproduction[6, 7]

According to a number of studies, the addition of a third dimension influences the psychological impact of viewers. This influence expresses itself in terms of experienced realism, power, and presence.[1] A more general higher level concept is needed to account for the added value of depth in 3D images. Therefore, we used the terms viewing experience and naturalness to quantify the added value of 3D images in respect to their 2D counterparts.

2. METHOD

2.1. Observers

Thirty observers were selected internally within Philips Research. Twenty observers participated in the viewing experience experiment and ten observers participated in the naturalness experiment. All participants had a visual acuity of ≥ 1, good stereo vision $<$30 seconds of arc (as tested with the Randot stereo test). The viewing distance was 1.5 meters.

2.2. Materials

2.2.1. Equipment

A 20" Philips multi-view autostereoscopic display was used in this experiment. The advantage of this display, besides 3D viewing without glasses, is the support of motion parallax enabling the viewer to look around objects by moving their head. Figure 1a shows an observer watching a set of objects. The left and right eye both receive a different view of the scene. By moving their heads, observers receive different views of the scene enabling them to see a potentially infinite number of views. Figure 1b shows the same viewing window, but this time divided into a finite set of horizontal frames. Each eye receives a view from a single frame, thereby preserving the effect of motion parallax, but with a reduced amount of views. Nine different views were generated using nine cameras (Figure 1c) and these nine views were integrated in the multi-view autostereoscopic display (Figure 1d). A set of nine successive views is called a viewing zone and repetition of this viewing zone enables multiple viewers to watch 3D. Figure 2 shows three zones consisting of nine views each. The resolution of the display was 1600x1200 pixels and the optics were optimized for a viewing distance of 1.5 meters. Custom build software was used to display the image material on the Philips multi-view autostereoscopic display.

2.2.2. Stimuli

The image material used in this experiment consisted of four still scenes, *Minibeamer*, *Puzzle*, *Rose* and *Shaver*, recorded with a nine camera set-up. The advantage of recording all the views with nine cameras instead of converting a 2D image into 9 views, is that all information is available and no depth distortions are visible in the 3D material. Displaying the nine views on the multi-view autostereoscopic display resulted in a 3D percept of the image because each eye receives a different view with a different perspective. The 2D situation was simulated by implementing the middle view (view five) into all nine views. In this case, the observer always perceives the same image on both eyes, resulting in a 2D percept. The middle view (camera five) of each image is shown in Figure 3.

In order to quantify viewing experience and naturalness, various image distortions could have been chosen. Artifacts like for instance blurring, blocking and ringing appear in different forms on different systems and their visibility depends on image content. Additive noise however seems to manifest itself in the same way over many different systems. Image independent noise can be described by an additive noise model (independent of image content), where the image $f(i,j)$ is the sum of the true image $s(i,j)$ and the noise $n(i,j)$. The model is shown in Equation 1.

$$f(i,j) = s(i,j) + n(i,j) \tag{1}$$

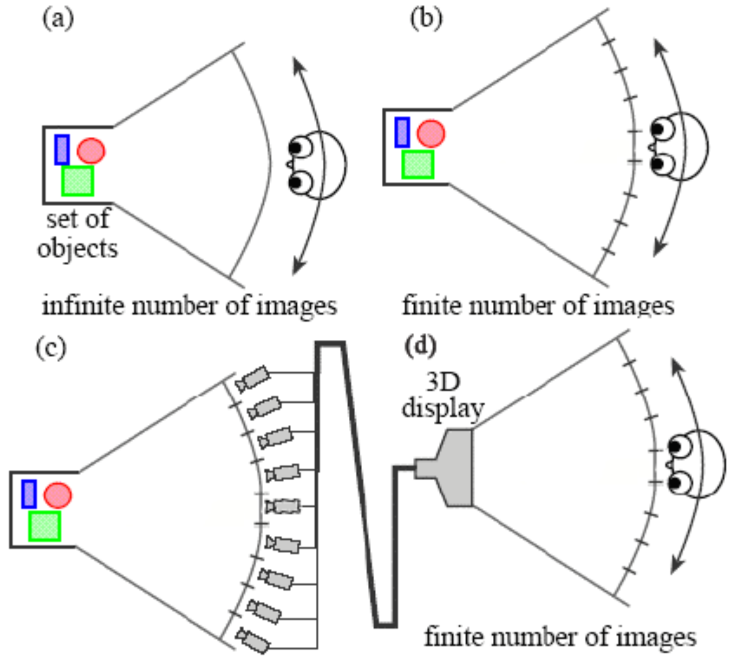
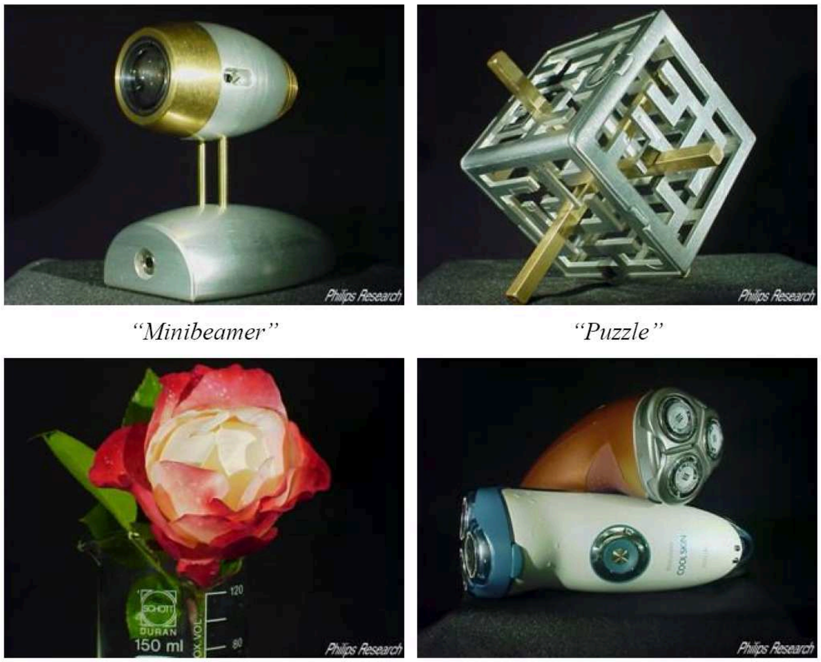


Figure 1. Panel (a) shows an observer watching a set of objects. The viewing window is divided in nine different perspective views in panel (b). In this experiment the nine different views were generated by using nine different cameras as shown in panel (c). The screen displays the nine different views in a viewing zone in panel (d).

Figure 2. Three viewing zones consisting of nine different perspective views each. The repetition of viewing zones enables multiple viewing.



"Minibeamer" *"Puzzle"*

"Rose" *"Shaver"*

Figure 3. The four panels show the original scenes *Minibeamer*, *Puzzle*, *Rose* and *Shaver*.

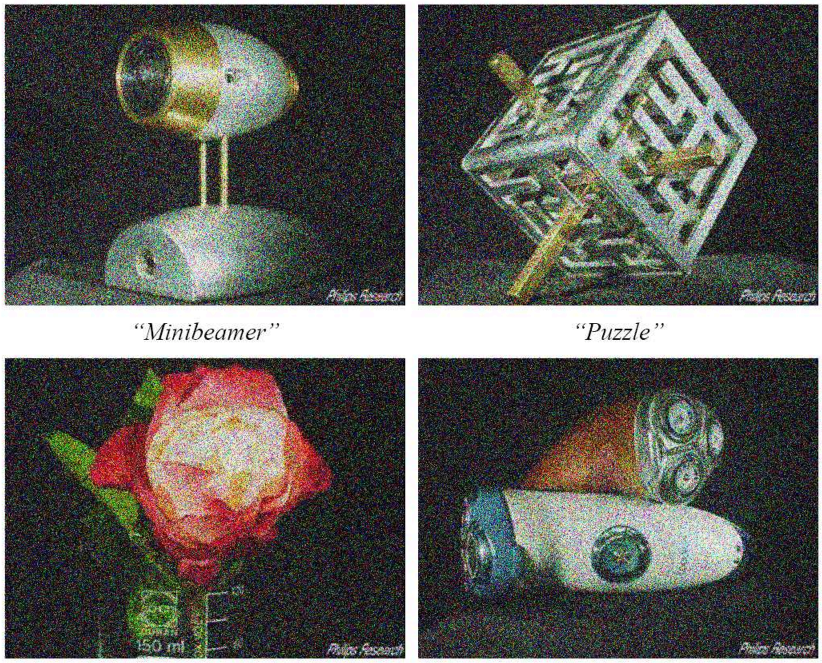
The noise is modeled with an independent, additive model, where the noise *n(i,j)* has a zero-mean ($\bar{x}$) Gaussian distribution described by its standard deviation (σ), or variance (σ^2). This means that each pixel in the noisy image is the sum of the true pixel value and a random, Gaussian distributed noise value. The additive noise is evenly distributed over the frequency domain (i.e. white noise). The white gaussian noise impairment was implemented using the Matlab image noise filter with five levels of noise ($\bar{x} = 0$, $\sigma^2 = 0.00125$, 0.0025, 0.005, 0.01, 0.02). An increasing σ^2-parameter produced more noise in the images. Figure 4 shows the four scenes with additive noise ($\bar{x} = 0$ and $\sigma^2 = 0.02$).

2.3. Procedure

The observers were given a brief introduction on paper about the experiment. Surfacing questions were answered and subsequently a short training session was conducted. The training session allowed the participants to get used to the setting as well as the tasks. In the training, six still images were presented with different noise levels, including the extremes used in the actual experiment. The rating scale for viewing experience and naturalness was labeled with the adjective terms [bad]-[poor]-[fair]-[good]-[excellent] according to the ITU[8] recommendation on subjective quality assessment. Participants were free to mark their assessment anywhere on the vertical rating scale. The order in which the images appeared was randomized throughout the experiment and each image was evaluated twice. In total, 20 participants had to indicate their viewing experience 96 times (4 images x 6 distortion levels (original + 5 noise impairment levels) x 2 conditions (2D and 3D) x 2 (rehearsal)). Exactly the same set-up was used for the naturalness ratings, only this session was done by 10 other participants. The images were displayed for 10 seconds followed by a grey field for 3 seconds. The lighting conditions of the room were constant for all participants and the level of light in the room was 25 lux, measured perpendicular to the display in the direction of the viewer.

3. RESULTS

Figure 5 shows the mean ratings for viewing experience averaged over the four images. On the x-axis the different noise levels are presented (increasing noise to the right). The y-axis represents the averaged values for viewing



"Minibeamer" *"Puzzle"*

"Rose" *"Shaver"*

Figure 4. Noise impaired scenes *Minibeamer*, *Puzzle*, *Rose* and *Shaver*.

experience from bad to excellent. The two lines in the figure represent the dimensions 2D and 3D. Error bars reflect the standard error of the mean.

A multivariate test (with noise, image and dimension as factors) was carried out on the raw subjective ratings to test the main effects and interactions for statistical significance. The results revealed a significant effect of image ($F(3,17)=6.413$, $p<.01$), dimension ($F(1,19)=5.251$, $p<.05$) and noise ($F(5,15)=46.521$, $p<.001$) on viewing experience ratings. No significant interactions between image, dimension and noise were found for any of the subjective ratings. Figure 5 clearly shows the main effect of noise on viewing experience ratings for both 2D and 3D images. The viewing experience of 3D images is rated higher than for 2D images for all noise levels explaining the main effect of dimension. The difference in viewing experience expressed in noise level is around 2 dB. Thus, 3D images with 2 dB more noise than their 2D counterparts result in equal viewing experience. So, the evaluation term viewing experience takes into account the added value of depth, as this is the only difference between the 2D and 3D images. The noise levels in 2D and 3D are perceptually the same.

Figure 6 shows the mean ratings for naturalness averaged over the four images. On the x-axis the different noise levels are presented (increasing noise to the right). The y-axis represents the averaged values for naturalness from bad to excellent. The two lines in the figure represent the dimensions 2D and 3D. Error bars reflect the standard error of the mean.

A multivariate test (with noise, image and dimension as factors) was carried out on the raw subjective ratings to test the main effects and interactions for statistical significance. The results revealed only significant effects of dimension ($F(1,19)=9.448$, $p<.013$) and noise ($F(5,15)=16.285$, $p<.004$) on naturalness ratings. No significant interactions between image, dimension and noise were found for any of the subjective ratings. Figure 6 clearly shows the main effect of noise on naturalness ratings for both 2D and 3D images. The naturalness of 3D images is rated higher than for 2D images for all noise levels explaining the main effect of dimension. The difference in naturalness expressed in noise level is around 4 dB.

Figure 5 and Figure 6 both show that noise considerably decreases viewing experience and naturalness ratings both for 2D and 3D. Furthermore, both figures show a higher score for the 3D-mode than the 2D-mode, which

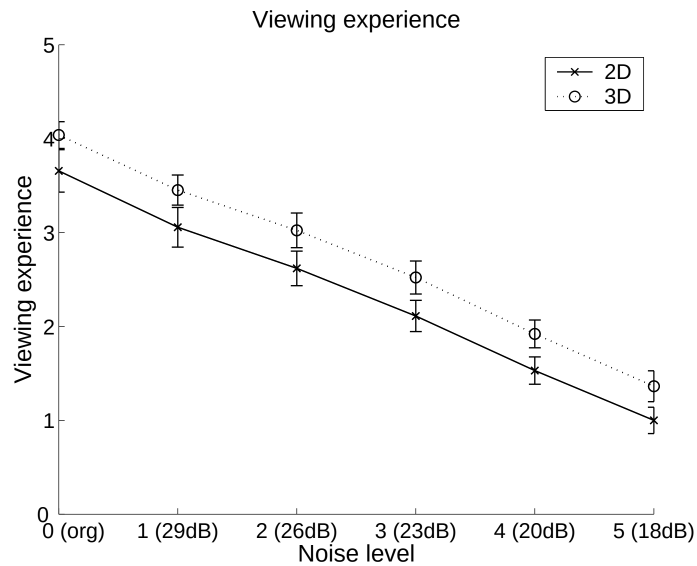


Figure 5. Viewing experience ratings averaged over all scenes. The x-axis represents the original image (org) and 5 noise impaired images (PSNR) and the y-axis represents the subjective ratings for viewing experience. The lines in the figure represent dimension (2D and 3D).

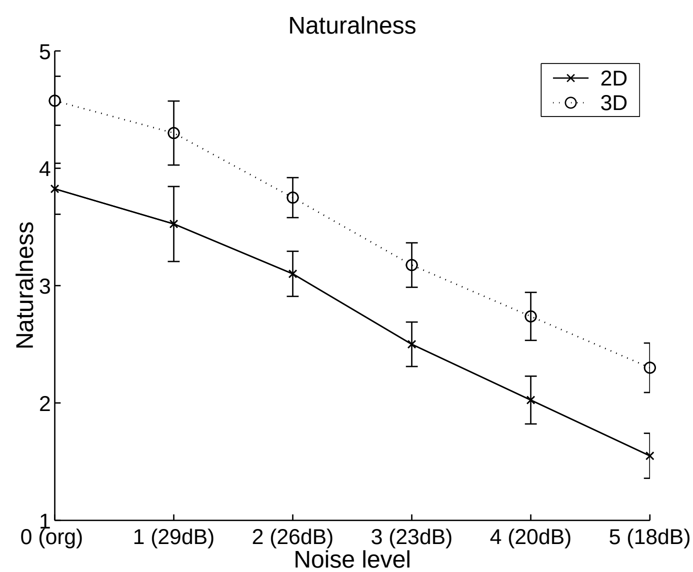


Figure 6. Naturalness ratings averaged over all scenes. The x-axis represents the original image (org) and 5 noise impaired images (PSNR) and the y-axis represents the subjective ratings for naturalness. The lines in the figure represent dimension (2D and 3D).

implies that both viewing experience and naturalness take into account the added value of depth. The difference between 2D and 3D is larger for naturalness than for viewing experience, which implies that the added value of depth is taken more into account in naturalness than in viewing experience. The fact that the difference between 2D and 3D ratings remains constant over all the noise levels implies that the perceived depth is independent of the noise level.

4. CONCLUSIONS

The results of the multivariate tests show that both noise and dimension significantly affect viewing experience and naturalness and there are no interaction effects. For viewing experience also image had a significant influence, but the added value of depth as measured by viewing experience was clearly recognized in all four images. Figure 5 and 6 show linear and parallel lines for both 2D and 3D ratings. Apparently, observers are well capable of assessing the added amount of noise to an image in the range used in this experiment. It is certainly possible that the plot will get the shape of a S-curve when extending the range of noise. At a first glance, the results show that observers anchor their judgments on the most salient features, in this case the noise level. However, by asking observers to assess viewing experience and naturalness, they do not only assess the induced noise level, but also other aspects in the image, which is illustrated by the fact that there are two distinctive lines for the assessment of 2D and 3D images. So, the added value of depth is taken into account with viewing experience, and even more with naturalness. Although present study demonstrated a significant difference in the assessment of 2D and 3D images, there are some questions still unanswered. Future research could determine to which extent the underlying aspects of viewing experience and naturalness are accountable for the difference in assessment of 2D and 3D images. Also more inside in the behavior of viewing experience and naturalness in combination with different 2D and 3D artefacts as well as moving 3D material will be needed.

5. ACKNOWLEDGMENTS

The authors would like to thank Philips Research Laboratories Eindhoven for the laboratory space and the technical support that made this project possible.

REFERENCES

1. W. IJsselsteijn, *Presence in Depth*, Eindhoven University of Technology, The Netherlands, 2004.
2. C. Smith and A. Dumbreck, “3-D TV: The practical requirements,” *Television: Journal of the Royal Television Society* , pp. 9–15, January/February 1988.
3. R. Sekuler and R. Blake, *Perception*, McGraw-Hill, New York, 2002.
4. P. Engeldrum, *Psychometric Scaling*, Imcotek Press, Winchester, Massachusetts, USA, 2000.
5. W. IJsselsteijn, H. de Ridder, and J. Vliegen, “Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain,” *Proceedings of the SPIE* **3957**, pp. 12–22, 2000.
6. W. Tam, L. Stelmach, and P. Corriveau, “Psychovisual aspects of viewing stereoscopic video sequences,” *Proceedings of the SPIE* **3295**, pp. 226–235, 1998.
7. P. Seuntiens, L. Meesters, and W. IJsselsteijn, “Perceptual evaluation of stereoscopic images,” *Proceedings of the SPIE* **5006**, pp. 215–226, 2003.
8. ITU, “Methodology for the subjective assessment of the quality of television pictures,” *Recommendation BT.500-10* , 2000.