Performance evaluation of 3D-TV systems

Ronald G. Kaptein[a], André Kuijsters[a,b], Marc T.M. Lambooij[a,b], Wijnand A. IJsselsteijn[b] and
Ingrid Heynderickx[a,c]

[a]Philips Research Laboratories, High Tech Campus 34, 5656AE, Eindhoven, The Netherlands
[b]Eindhoven University of Technology, P.O. Box 513, 5600MB, Eindhoven, The Netherlands
[c]Delft University of Technology, Mekelweg 4, 2628CD, Delft, The Netherlands

ABSTRACT

The image quality circle is a commonly accepted framework to model the relation between the technology variables of a display and the resulting image quality. 3D-TV systems, however, go beyond the concept of image quality. Research has shown that, although 3D scenes are clearly more appreciated by subjects, the concept 'image quality' does not take this added value of depth into account. Concepts as 'naturalness' and 'viewing experience' have turned out to be more useful when assessing the overall performance of 3D displays. In this paper, experiments are described that test 'perceived depth', 'perceived image quality' and 'perceived naturalness' in images with different levels of blur and different depth levels. Results show that naturalness incorporates both blur level as well as depth level, while image quality does not include depth level. These results confirm that image quality is not a good measure to assess the overall performance of 3D displays. Naturalness is a more promising concept.

Keywords: 3D, image quality, viewing experience, naturalness

1. INTRODUCTION

The major evaluation criterion to assess the subjective performance of an imaging system is perceived image quality. Engeldrum[1] defines image quality as "the integrated set of perceptions of the overall degree of excellence of the image". Of main interest for display manufacturers is in general the relation between the technology parameters of the imaging system and the perceived image quality. Examples of technology parameters are pixel pitch, type of color filters and video-processing parameters. However, it is often very time consuming and inefficient to assess the relation between image quality and technology variables directly, especially because this process has to be repeated for any new display design again, since the optimization does not allow generalization. To overcome this, Engeldrum[1,2] has developed a model to evaluate this relation in another way. His Image Quality Circle breaks the relation between technology parameters and image quality into measurable steps. As can be seen in Fig. 1, he first relates image quality to so-called image-quality attributes (step 2 in Fig. 1), using a so-called image-quality model. Image-quality attributes are characteristics of the image that are defined at the perception level of the viewer and are unconsciously weighted to come to an overall image quality judgment. Examples are brightness, colorfulness, sharpness, and also noise and ringing. These attributes can be linked to the so called physical characteristics of the image (step 3 in Fig. 1). This can be done through visual algorithms. Physical image characteristics can be physically measured, and as a consequence, have physical dimension units. Examples are chromaticity of the display primaries, the light intensity or the gamma. In the end, systems models can be used to predict these physical image characteristics from the technology variables (step 4 in Fig. 1). If we have a good understanding of steps 2 to 4, it is in principle relatively easy to determine the image quality of a new or hypothetical system.

Image quality has proven to be a good method to evaluate the subjective performance of 2D imaging systems, and the image-quality circle is a reliable approach for this. However, the visual impact of recent display enhancements, such as Ambilight or 3D, goes beyond the concept of image quality.[3-7] Image quality is still an important aspect of the subjective performance of the display, but it is not all-embracing anymore. To fully

---

Further author information: (Send correspondence to R.G. Kaptein)
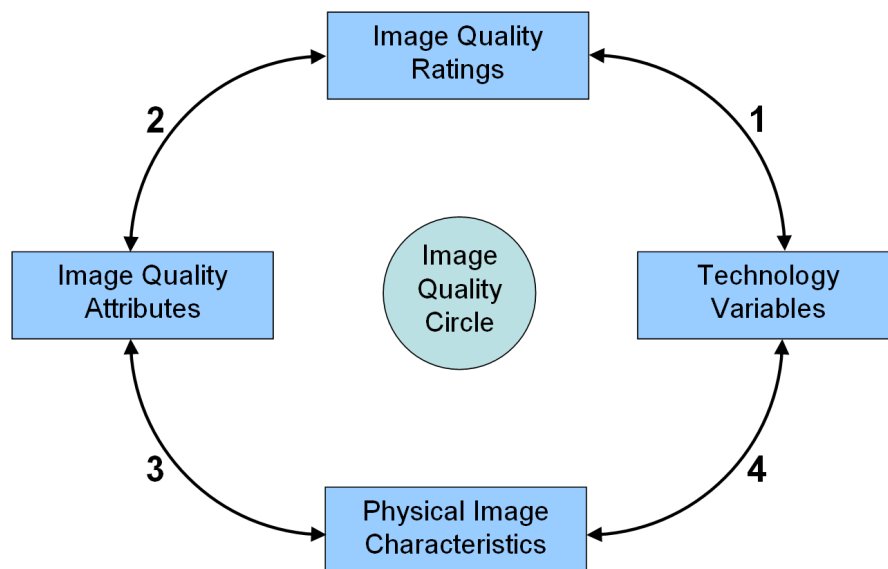R.G. Kaptein: E-mail: ronald.kaptein@philips.com, Telephone: +31 (0)40 2747904

Figure 1. The image quality circle of Engeldrum,[1,2] which relates image quality to technology variables.

grasp the perceptual performance of these new systems, the Image Quality Circle needs expansion. This paper focuses on the expansion of the model for 3D-TV.

2. THEORY

2.1. 3D DISPLAYS

Our visual system constructs a 3D representation of the outside world. To do this, it utilizes various visual cues. These cues can be divided into four categories[8]: ocular information (accommodation and convergence), dynamic information (motion parallax), pictorial information (e.g. perspective, occlusion) and stereoscopic information (binocular disparity). A lot of these cues are monocular, meaning that they are visible with one eye. One of the major sources of depth perception is binocular disparity. It refers to the fact that both eyes receive slightly different views of the outside world. Based on this disparity, the brain is able to create the impression of depth. This process is called stereopsis.

Binocular disparity can be used in display applications to create a depth impression. The challenge is then to present different images (also called views) to the left and right eye. Already for a long time this can be accomplished using special glasses (e.g. polarized or shutter glasses). Nowadays also solutions that do not need glasses are available, so called auto-stereoscopic displays. The two main types of auto-stereoscopic displays are barrier displays and lenticular displays. In barrier displays, a barrier structure in front of the pixels blocks the light from certain pixels for one eye, but not for the other eye. In this way, two (or more) different views can be created. The drawback of this method is that light blocked by the barrier is lost, so the luminance of a barrier display is typically quite low. In lenticular displays, a lenticular lens sheet is put in front of the pixels. The lenses direct the light from different pixels into different directions, creating different views. Philips, for example, has created an auto-stereoscopic lenticular 3D display with 9 views.[9,10] Fig. 2 illustrates the basic principle of this display. Note that, if more than two views are created, the display has the possibility of creating motion parallax and look-around capabilities.

2.2. PERFORMANCE OF 3D TVs

The depth introduced in 3D-TVs clearly enhances the overall subjective performance of the TV. This is acknowledged by virtually all people watching a 3D-TV. Unimpaired 3D images are more appreciated than their 2D
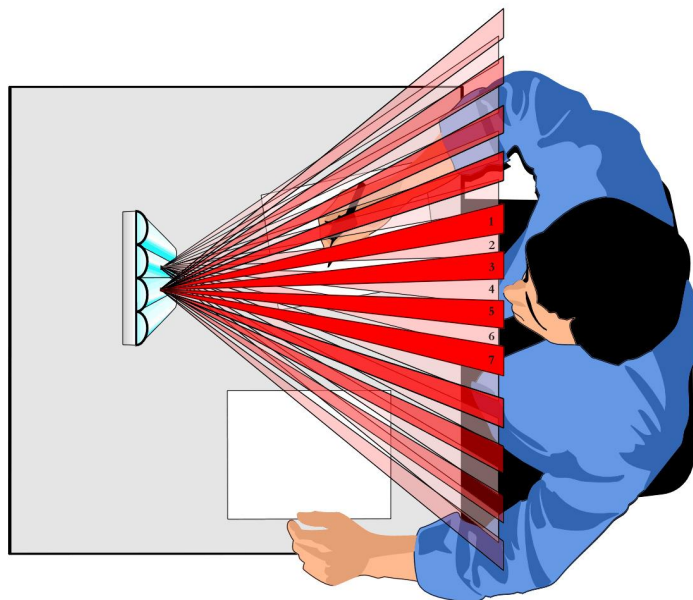
Figure 2. Philips auto-stereoscopic 3D display uses a lenticular to provide the viewer with multiple views.

counterparts. So, to assess the overall performance of a 3D-TV, depth is something that should be included in the used evaluation criterion. In other words, a subjective measure or criterion needs to be found that is judged to be higher for 3D material than for otherwise similar 2D material.

Since image quality has proven to be a reliable way to assess the performance of 2D imaging systems, it is obvious that this concept is also applied to 3D systems. However, it has already been shown that the added value of perceived depth is not incorporated when subjects assess image quality.[4, 6, 7, 11] In one of these studies,[11] perceived image quality is determined for various JPEG-compression levels and various camera-base distances (i.e. depth levels). Fig. 3 shows this data, and it is very clear that image quality decreases with increasing compression, but is independent of depth level.

This means that another concept for the overall performance of 3D displays is needed. "Presence", "naturalness" and "viewing experience" have been mentioned as possibilities for such a concept.[8] The concept of presence has been described as "the subjective experience of being in one place or environment, even when one is physically situated in another".[12] Naturalness refers to the degree in which the image approximates reality. Viewing experience is a complex, highly multidimensional, high-level concept and cannot easily be defined. One could say that in the imaging industry, "enhancing the users' viewing experience" is in the end what it is all about.

It has been shown that there is a correlation between image quality and presence.[7, 13, 14] It has also been shown that perceived depth increases the feeling of presence.[15, 16] However, this effect is probably quite small.[17] It has been shown that presence is a suitable attribute for evaluating stereoscopic video sequences, but it is also suggested that it maybe less appropriate for stills.[7, 17]

Previous research showed a correlation between image quality and naturalness,[18, 19] but also that these concepts are not identical.[19, 20] Also a correlation between depth and naturalness has been shown.[6, 16] Seuntiëns et al[6] measured perceived naturalness as a function of noise level, for 2D and 3D still images. The results, shown in Fig. 4, clearly show that naturalness decreases with noise level, both for 2D and 3D material, and that, for a given noise level, 3D material is perceived as significantly more natural than 2D material. In a more recent study, Seuntiëns et al,[7] using video sequences, again showed that there is an effect of depth on naturalness,
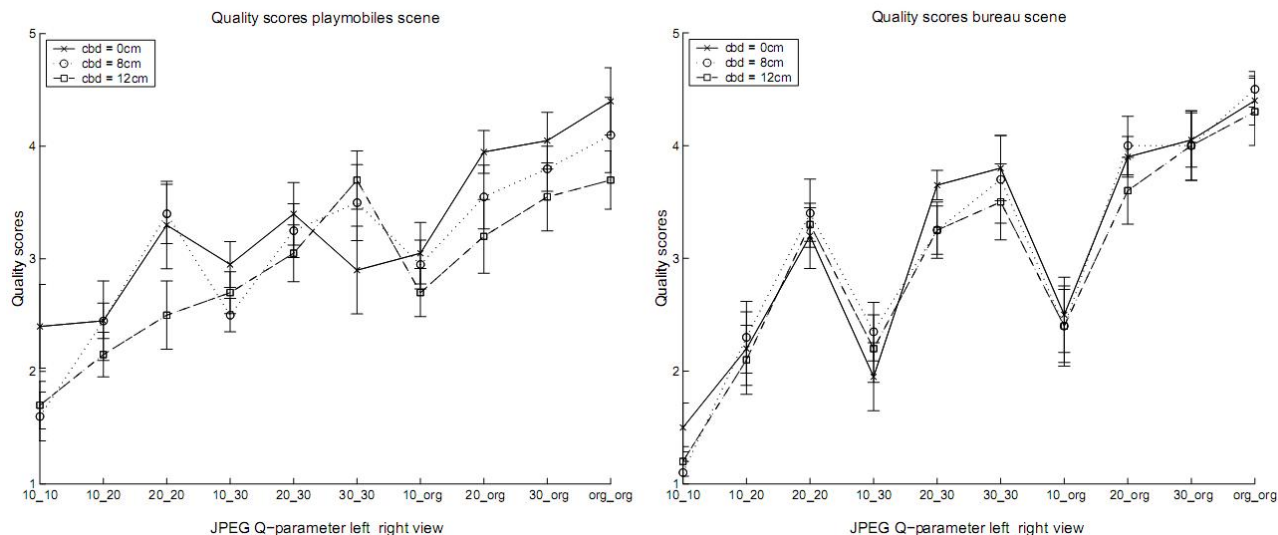
Figure 3. Image quality scores as a function of JPEG Q-parameter (average bit rate, which could be different for left and right eye, increases to the right), for the two scenes used in the experiment. The three different lines indicate three different camera-base distances (i.e. depth levels). It can be seen that image quality on average increases with bit rate, but is independent of depth level. Taken from Seuntiens et al.[11]

but this effect is less pronounced. That study also showed that increasing depth too much actually decreases naturalness due to the artifacts that are created.

Besides naturalness, Seuntiëns et al[6] also measured perceived viewing experience as a function of noise level, for 2D and 3D still images. The results (Fig. 5) are similar to the naturalness results (Fig. 4). Viewing experience decreases with noise level, both for 2D and 3D material. For a given noise level, 3D material leads to a higher viewing experience than 2D material. The recent study of Seuntiëns et al[7] also showed an effect of depth on viewing experience, demonstrating that increasing depth too much decreases viewing experience due to the occurrence of 3D artifacts.

It has thus been shown that presence, naturalness and viewing experience take both depth and image quality into account, but not all to the same extent. It has been suggested that presence is mainly suitable for video sequences, and less for stills. This makes presence less suitable as a new concept to assess the overall performance of 3D material. In Seuntiëns et al (2005)[6] naturalness seemed to incorporate depth more than viewing experience (see also Fig. 4 and 5), making it a little more appropriate to assess 3D performance. Recently, Seuntiëns et al (2007)[7] found somewhat stronger effects of viewing experience and presence compared to naturalness, but he used video sequences instead of stills.

To make things a little bit more complicated, including depth may also have a negative contribution to the image quality, and as such to the preferred presence, naturalness or viewing experience. Current autostereoscopic displays unavoidably have a lower resolution than a typical 2D display, and the different views are never fully separated in space (a phenomenon known as crosstalk[21]) which leads to blurring and ghosting. These "3D quality" aspects negatively influence image quality.[22] Another issue is that too much depth may lead to visual discomfort or eye strain. It has indeed been shown that ratings of visual strain increase with increasing depth.[22, 23]

Until now, naturalness and viewing experience have been found to include both the effect of image quality as well as the added value of depth. However, this was only tested for a limited set of attributes. It has been shown that for crosstalk (i.e. blur and ghosting) and JPEG-quality, image quality is not the right measure, but other concepts were not tested. Only for noise[6] and bitrate[7] the concepts of naturalness and viewing experience were more or less validated. In order to end up with a solid new measure, these concepts need to be tested more
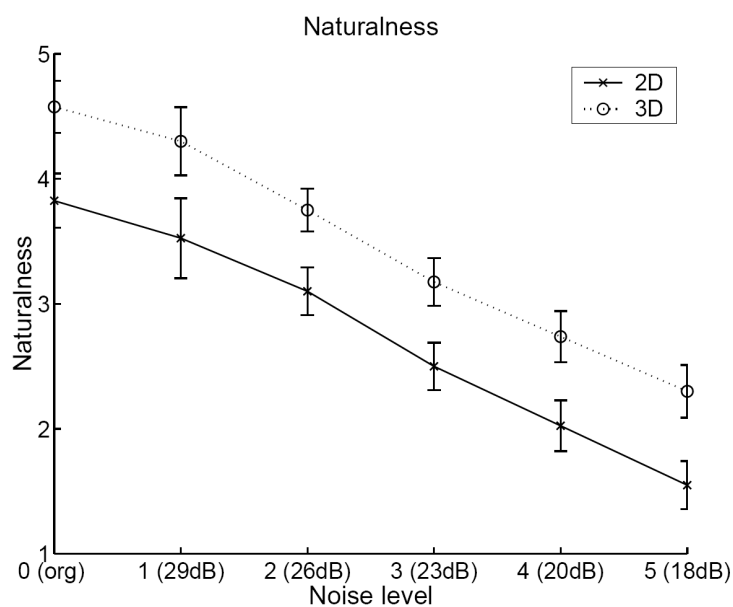
Naturalness

Figure 4. Naturalness as a function of noise level, for 2D (solid line) and 3D (dashed line) images. Naturalness decreases with noise level, but is clearly higher for 3D than for 2D images. Taken from Seuntiëns et al.[6]
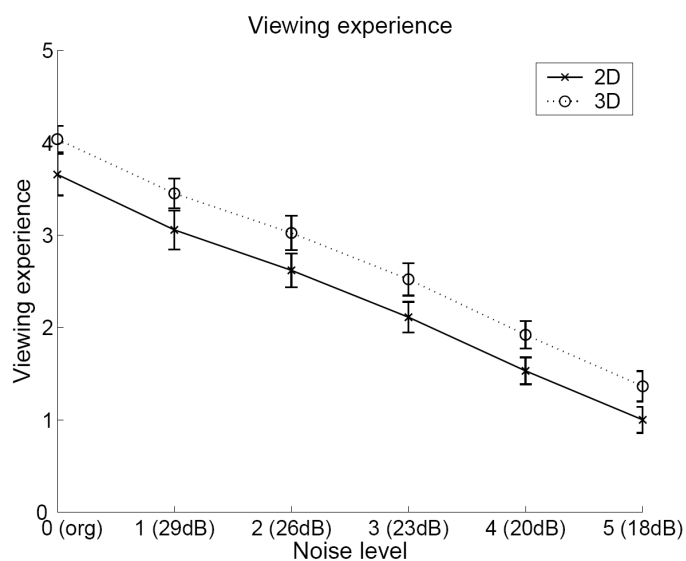
Viewing experience

Figure 5. Viewing experience as a function of noise level, for 2D (solid line) and 3D (dashed line) images. Viewing experience decreases with noise level, but is somewhat higher for 3D than for 2D images. Taken from Seuntiëns et al.[6]

Figure 6. Scenes used in the experiment: playmobiles (top left), puzzle (top right), nature (bottom left) and balloon (bottom right).

thoroughly. In this paper, experiments assessing perceived depth, image quality and naturalness in blur-impaired images are presented, and the implications for a new 3D viewing-experience model are discussed.

3. EXPERIMENTAL SETUP

The stimuli were displayed on a 42" Philips nine-view auto-stereoscopic lenticular display. The viewing distance was 3 m. Four still images were used in the experiment (see Fig. 6). The input format for the Philips 3D display was an RGB image (original) + corresponding Z-image (depth map), with a resolution of 940 x 540 each. The depth map was a gray-scale image in which the gray value indicated the relative depth of each pixel (255 is maximal towards the viewer, 0 is maximal away from the viewer). Three different depth levels were created by varying the gain factor of the depth map. They resulted in a 2D level (zero screen disparity) and two 3D levels corresponding to moderate (3Dhalf) and high (3Dfull) depth (the increase in depth (i.e. disparity) between 2D and 3Dhalf was the same as between 3Dhalf and 3Dfull). Four levels of blur were introduced to each scene, using a Gaussian filter with a kernel size of 7 x 7 pixels and a standard deviation (σ) of 0, 1, 1.5 and 5.

Nineteen observers participated in the experiment. Ten of them were employees of Philips Research, while nine were internal graduation students. All observers had a good visual acuity (≥ 1 as tested with the Landolt-C test) and good stereovision (≤ 30 second of arc as tested with the RANDOT stereo test). The subjects were requested to score the perceived depth, perceived image quality and perceived naturalness for all 48 stimuli (i.e. 4 different scenes x 4 levels of blur x 3 depth levels). Each attribute was assessed in a different session. Subjects judged all images using a scale which was labeled with the adjectives [bad]-[poor]-[fair]-[good]-[excellent].[24] Between these five standard marks four intermediate marks were added to allow the participants to refine their assessment, which resulted in a nine point scale.

Because the attributes were rated on an ordinal categorical scale, the data were first transformed using Thurstone scaling, more particularly with the ThurcatD software.[25] With this scaling it was tested whether equal distances on the assessment scale were perceived as equal, e.g. whether the difference between [bad] and [poor] was perceived similar as the difference between for example [fair] and [good]. The resulting data were analyzed in SPSS 15.0. In a repeated measure ANOVA the main effects of blur level, depth level and scene, as well as all three two-way interactions, were tested for each assessment criterion separately, i.e. for perceived depth, perceived image quality and perceived naturalness. The Bonferroni post hoc test was used to specify significant main effects and their interactions.

4. RESULTS

The ThurcatD analysis shows that the steps on the assessment scale are also perceptually equal. It is therefore allowed to perform the analysis on the raw data instead of on the ThurcatD results. A first analysis of the data reveals that the 'puzzle' scene is scored very different compared to the other scenes: image quality and naturalness are scored considerably lower for the two 3D settings. The ANOVA analysis reveals that when the 'puzzle' scene is included, depth has a significant (negative) effect on image quality, whereas this effect is absent when the 'puzzle' scene is removed from the data set. This is most likely due to the fact that parts of the 'puzzle' scene, because of its properties, become very blurred due to the crosstalk between the views,[21] resulting in an interaction between blur level and depth level. This interaction effect is to a much lesser extent, if at all, present in the other three scenes. Based on this, we decided to remove the 'puzzle' scene from the pooled data set and to discuss it separately. For the remaining data set, there is still a significant effect of scene ($p < 0.001$ for all 3 attributes). In the statistical analysis, the differences between the scenes are taken into account, but since all trends and effects are in the same direction, only the averaged results are shown in the figures.

Fig. 7 summarizes the averaged results. It is clear that the scores of all three attributes decrease with increasing blur. This is very obvious for the image quality and naturalness score, but it is also true for the depth scores, albeit to a lesser extent.

Figure 7 also demonstrates that perceived depth increases with screen disparity (i.e. depth level: 2D, 3D half and 3D full), which is of course as expected. This effect is significant ($F(2, 34) = 155.102; p < 0.001$). There is also a significant decrease of perceived depth with increasing blur ($F(3, 51) = 27.791; p < 0.001$). This effect is similar for each blur level. In other words: introduced blur degrades the depth perception of each depth level in a similar way. The increase of perceived depth with depth setting is thus independent of blur.

The center panel of Fig. 7 shows that there is a strong effect of blur on image quality ($F(3, 42) = 446.872; p < 0.001$). However, the figure also clearly shows that depth does not significantly influence perceived image quality, which was confirmed by the statistical analysis ($p = 0.07$). The bottom panel of Fig. 7 shows that the naturalness scores are also strongly influenced by the amount of blur ($F(3, 48) = 146.481; p < 0.001$). Naturalness is also significantly higher for the 3D levels than for the 2D level ($F(2, 32) = 13.496; p < 0.001$). However, the naturalness scores are the same for the two depth levels (3D half and 3D full). So, naturalness takes the added value of depth into account, but only the introduction of depth. Further increasing depth does not increase the naturalness scores.

The ANOVA analysis also reveals some interaction effects (depth x scene for perceived depth and perceived image quality, blur x scene for perceived image quality and perceived naturalness). These suggest that different scenes are affected differently by introduced blur, and also by introduced depth. This is of course not really surprising.

The 'puzzle' scene was removed from the main analysis, and is briefly discussed here separately. There is a large interaction between depth level and blur level in this scene, and an ANOVA analysis indeed reveals a significant effect of depth on image quality ($F(2, 32) = 37.483, p < .001$) for this specific image. Also the naturalness scores are affected by the depth level; the 3Dfull naturalness scores are significantly lower than the two other depth levels. There is no significant difference between 2D and 3Dhalf. Since the image-quality scores are significantly lower for the 3Dhalf depth level compared to the 2D depth level ($p < 0.01$), this suggest that the added value of depth is taken into account, despite the fact that 3D artifacts are present.

5. DISCUSSION

The results presented in this paper again confirm that the criterion image quality does not take the added value of depth into account. They also show that naturalness incorporates both blur as well as depth level. This supports the notion that naturalness is a valid concept to evaluate the overall visual performance of 3D displays, in the sense that it includes both the negative aspect of image blur as well as the positive aspect of depth. As a result, it confirms the earlier hypothesis that naturalness is a weighted average of image quality and depth.[26] This leads to the simple conceptual model depicted in Fig. 8. Note that for perceived depth, a concept similar to the image-quality circle could be developed, relating perceived depth to technology variables, via depth attributes and physical image characteristics. For example, the present finding that perceived depth is dependent on the
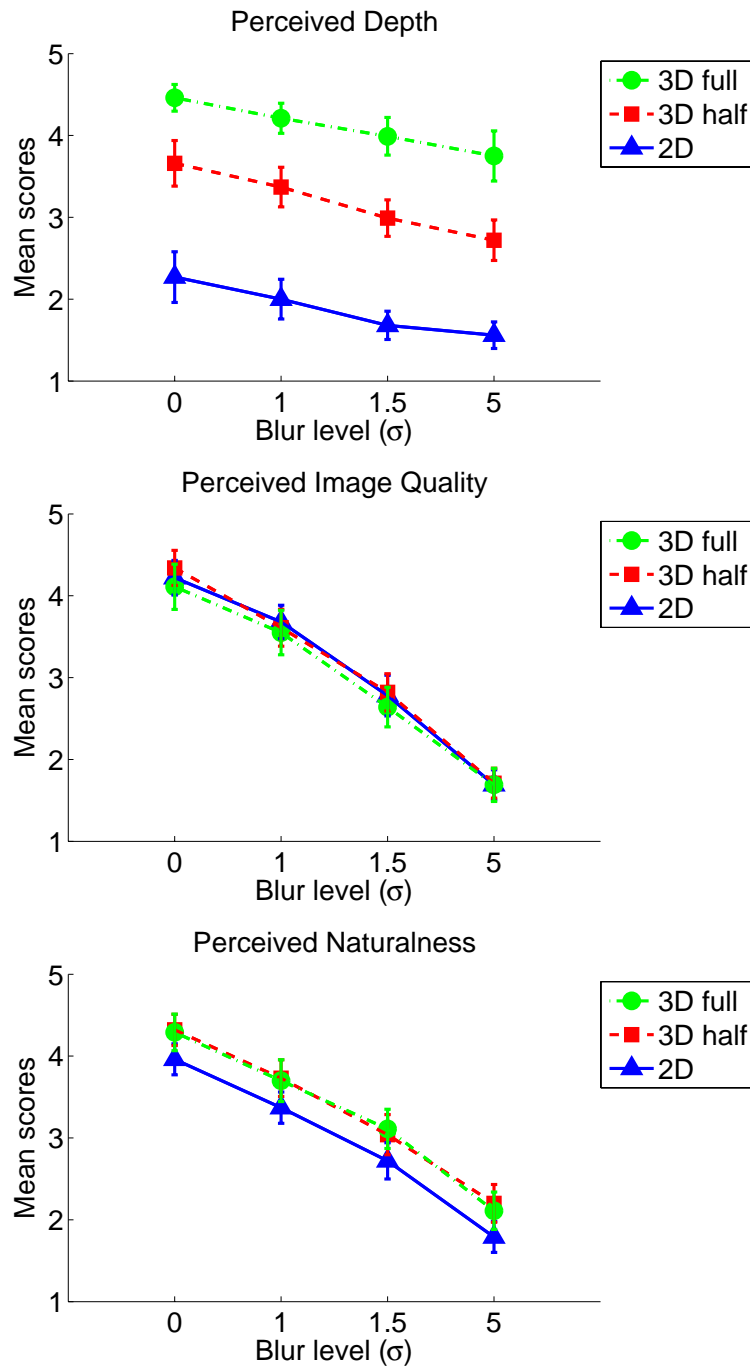
Figure 7. Mean assessment scores for perceived depth, image quality and naturalness, as a function of blur level. The three lines represent the different depth levels. Error bars show 95% confidence intervals.

amount of blur suggest that depth is not merely a function of the binocular disparity, but that other attributes may also play a role. However, at present too little is known about the attributes that construct perceived depth to create such a 'depth circle'.

The conceptual model seen in Fig. 8 could be implemented quantitatively by doing a linear regression of
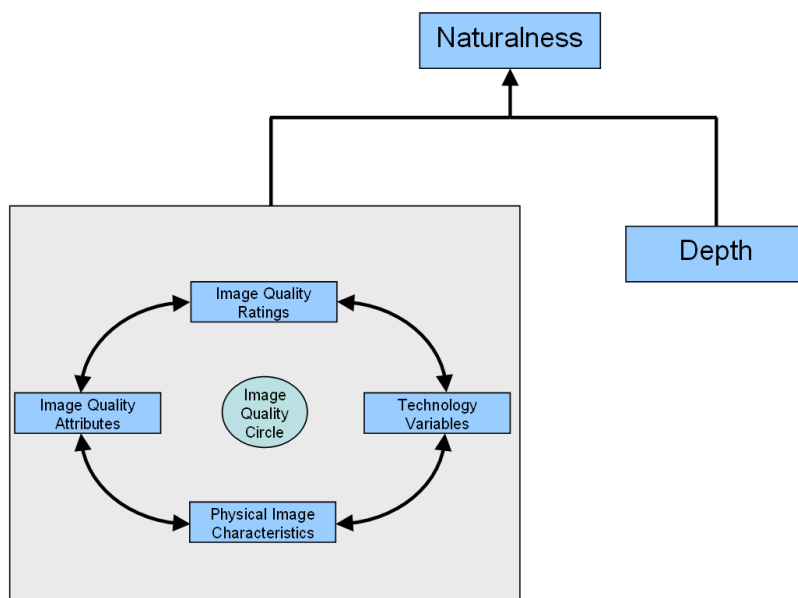
Figure 8. Model proposed for the evaluation of the visual performance of a 3D display

naturalness scores on depth scores and image-quality scores.[26] This way the relative weights of depth and image quality could be determined. However, the current results are not yet suitable to construct such a quantitative model, since the scales of the three attributes are relative. This is due to the fact that subjects were asked to use the full scoring range of 1 to 5 for each attribute, while the perceptual ranges of the three attributes are most likely not the same. Future research will have to deal with this problem to make such a model possible.

As stated earlier, image quality itself can be negatively affected by stereoscopic depth. This has been shown in earlier research[22] and it is probably also the case for the 'puzzle' scene of the current data. It may also have played a role for the other scenes, in the sense that the difference in naturalness scores between 2D and 3D may have been bigger if the 3D artifacts would not be present. This relation between image quality and depth is not evident from Fig. 8. However, there is a reason for that. The effect of stereoscopic depth (i.e. display disparity) on image quality is essentially a 2D artifact[22] and has nothing to do with the concept of perceived depth itself, which is the reason that it is not explicitly included in Fig. 8. In other words, the binocular disparity in a 3D display leads to perceived depth and to artifacts such as blur, that are incorporated in image quality. So, both are two fundamentally different concepts. This differentiation was already acknowledged by Seuntiëns et al.[22] Another argument supporting this is that these artifacts are strongly, or even mainly depending on the type of system used. In autostereoscopic displays, there is typically a lot of crosstalk and therefore blur, while in other 3D systems (such as a Wheatstone viewer[22]), crosstalk can be completely absent. In other words, in future 3D displays the (apparent) link between image quality and depth may be resolved.

As stated earlier, another factor that has an effect on the visual performance of 3D displays is visual (dis)comfort. In 3D displays, visual comfort can be drastically reduced through for example binocular asymmetry[27] or possibly the accommodation-vergence conflict.[28] How visual comfort plays a role in the overall viewing experience is not yet known. Seuntiëns et al[22] proposed a conceptual model including image distortion, perceived depth and visual strain, in which each of these factors affects viewing experience. Whether visual comfort acts for example on perceived naturalness or on a higher concept is not known. Is should also be noted that visual comfort is currently mainly assessed through subjective evaluation (e.g. questionnaires), which is very informative, but having a more objective method next to that would be very helpful in order to include visual comfort in the model in a reliable way.

The main conclusion from these new results is that 'perceived naturalness' is a good concept to assess the

overall performance of 3D displays, and that we are a small step closer to the construction of a '3D visual experience' model and as a result also to an 'overall visual experience' model. This latter model will potentially bear relevance to a broad range of visual experiences, such as those induced by, for example, Ambilight displays.[7]

REFERENCES

1. P. Engeldrum, "Image quality modeling: Where are we?," in *PICS Conference Proceedings*, 1999.
2. P. Engeldrum, "A theory of image quality: The image quality circle," *Journal of Imaging Science and Technology* **48**, pp. 447–457, Jan. 2004.
3. W. Tam, L. Stelmach, and P. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," in *Proceedings of the SPIE*, **3295**, pp. 226–235, SPIE, 1998.
4. P. Seuntiëns, L. Meesters, and W. IJsselsteijn, "Perceptual evaluation of JPEG-coded stereoscopic images," *Proceedings of SPIE* **5006**, p. 215, 2003.
5. L. Meesters, W. IJsselsteijn, and P. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology* **14**(3), pp. 381–391, 2004.
6. P. Seuntiëns, I. Heynderickx, and W. IJsselsteijn, "Viewing experience and naturalness of 3D images," in *Proceedings of the SPIE*, 2005.
7. P. Seuntiëns, I. Vogels, and A. van Keersop, "Visual experience of 3D-TV with pixelated ambilight," in *Proceedings of PRESENCE 2007*, 2007.
8. P. Seuntiëns, *Visual experience of 3D TV*. PhD thesis, Eindhoven University of Technology, 2006.
9. C. van Berkel and J. Clarke, "Characterization and optimization of 3D-LCD module design," in *Proceedings of the SPIE*, *Proceedings of the SPIE* **3012**, p. 179, SPIE, 1997.
10. O. H. Willemsen, S. T. D. Zwart, and M. G. H. Hiddink, "2-D/3-D switchable displays," *Journal of the Society for Information Display* **14**, pp. 715–722, 2006.
11. P. Seuntiëns, L. Meesters, and W. IJjsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Transactions on Applied Perception* **3**, pp. 95–109, Jan. 2006.
12. B. Witmer and M. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and Virtual Environments* **7**, pp. 225–240, 1998.
13. W. Barfield and S. Weghorst, "The sense of presence within virtual environments: A conceptual framework," *Proceedings of the Fifth International Conference on Human-Computer Interaction* **2**, pp. 699–704, 1993.
14. C. Bracken, "Presence and image quality: The case of high-definition television," *Media Psychology* **7**(2), pp. 191–205, 2005.
15. J. Freeman and S. Avons, "Focus group exploration of presence through advanced broadcast services," in *Proceedings of the SPIE*, **3959**, pp. 530–539, 2000.
16. W. IJsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman, "Perceived depth and the feeling of presence in 3DTV," *Displays* **18**(4), pp. 207–214, 1998.
17. W. IJsselsteijn, H. de Ridder, J. Freeman, S. Avons, and D. Bouwhuis, "Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence," *Presence: Teleoperators and Virtual Environments* **10**(3), pp. 298–311, 2001.
18. E. Fedorovskaya, H. de Ridder, and F. Blommaert, "Chroma variations and perceived quality of color images of natural scenes," *Color research and application* **22**(2), pp. 96–110, 1997.
19. H. de Ridder, "Naturalness and image quality: saturation and lightness variation in color images of natural scenes," *Journal of Imaging Science and Technology* **40**(6), pp. 487–493, 1996.
20. H. de Ridder, F. Blommaert, and E. Fedorovskaya, "Naturalness and image quality: chroma and hue variation in color images of natural scenes," *Proceedings of the SPIE* **2411**, pp. 51–61, 1995.
21. R. Kaptein and I. Heynderickx, "32.2: Effect of crosstalk in multi-view autostereoscopic 3D displays on perceived image quality," *SID Symposium Digest of Technical Papers* **38**, pp. 1220–1223, 2007.
22. P. J. H. Seuntiëns, L. M. J. Meesters, and W. A. IJsselsteijn, "Perceptual attributes of crosstalk in 3D images," *Displays* **26**, pp. 177–183, Oct. 2005.

23. W. IJsselsteijn, H. de Ridder, and J. Vliegen, "Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain," in *Proceedings of the SPIE*, **3957**, pp. 12–22, SPIE, 2000.

24. ITU, "500-11. methodology for the subjective assessment of the quality of television pictures," *RECOMMENDATION ITU-R* **BT.500-11**, 2002.

25. M. C. Boschman, "ThurCatD: A tool for analyzing ratings on an ordinal category scale," *Behavior Research Methods, Instruments, & Computers* **32**, pp. 379–388, 2000.

26. I. Heynderickx, "Performance evaluation of 2D and 3D-TV systems," in *Int. Congress of Imaging Science*, (Rochester, New York), 2006.

27. F. L. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays* **25**, pp. 99–108, Aug. 2004.

28. M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *Journal of Display Technology* **1**(2), pp. 328–340, 2005.