Global Brightness Contrast and the Effect on Perceptual Image Quality

Jacques A.J. Roufs
Victor J.F. Koselka
Arthur A.A.M. van Tongeren

Institute for Perception Research (IPO)
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

The perceptual image quality of natural scenes as a function of the physical system parameter gamma has a definite optimum. This optimum is subject-independent and greater than 1, but was found to vary from one scene to another. If gamma is varied, brightness contrast is the most obviously changing perceptual attribute. Subjects appear to be able to make consistent, global judgements of brightness contrast in natural scenes, despite the fact that local brightness contrast may vary considerably. If scaled perceptual quality is plotted against scaled (perceived) brightness contrast, all curves coincide, suggesting that under the given conditions brightness contrast is the dominant psychological dimension of the perceptual image quality. Taking into account the grey-level distribution of the scene in combination with the luminance-reproduction function of the imaging chain, an effective gamma value can be defined. If scaled perceptual quality and global brightness contrast are plotted against this effective gamma, the differences between scenes disappear, although there are clear differences in the relative sizes of the light and dark parts of the various test scenes. An analysis by scaling global brightness of gaussian blobs of randomly distributed sizes, modulation depths and polarities, shows that skewness of this distribution does indeed have only a weak effect over a considerable range. The data suggest that the ratio of maximum and minimum luminance determines global brightness contrast for complex scenes under these conditions.

Introduction

The fast developments in display technology and image technology in general have caused a growing interest in the fundamental aspects of perceptual image quality. Perceptual image quality expresses a degree of excellence of the image. Although it is hard to define it exactly, subjects are able to order this complex psychological attribute fairly consistently with respect to its strength. It has been shown, for instance, that images scaled by groups of subjects in different countries, using adjectives as categories, match quite well.[1] Understanding of the relationship between perceptual image quality, its underlying sensorial-attributes and the physical image parameters is quite important for optimizing display designs or digital coding or processing algorithms.

In this paper we will focus attention on brightness contrast, which is generally recognized as being one of the important factors in perceptual image quality (see overview[2]). Brightness contrast of complex patterns is not a simple sensorial attribute. It is not yet fully understood how brightness contrast is related to the brightness pattern. Firstly, although brightness contrast is sometimes associated with the differences or the ratio of local brightnesses the true relationship is still a matter of debate. Secondly, it is not known what effect the size of the areas with different brightnesses has on the contrast impression. One intuitively doubts whether, for example, a tiny but very dark spot in the scene, caused by a particle of dust, would seriously affect the brightness contrast of a picture, while a large dark patch in the scene certainly will. Thirdly, in the case of complex natural images, subjects also have an impression of global contrast, connected with the pattern of brightnesses as a whole. The strength of this global perceptual attribute can be judged in a reliable way and its impression arises apparently by integrating local brightness contrasts in some unknown way. Fortunately, in the context of perceptual image quality we do not need to solve these problems, if we only want to measure the strength of this global psychological attribute properly. Experience has shown that subjects manage to judge global brightness in a stable, repeatable way, with no alarming differences between individuals.[3,4]

While investigating global brightness contrast there is the problem of finding an adequate physical parameter. In the world of display engineers, ratios of luminances are apparently preferred in simple cases, like the two-level configuration of visual display terminals (VDTs). In more complex situations, like ordinary TV pictures, where many grey levels are involved, other contrast measures have to be considered. Frequently used measures are the ratio of maximum to minimum luminance in the scene or, taking into account all levels between, the slope of a linear luminance-reproduction curve. With respect to the latter, however, the luminance-reproduction function of TV chains is in fact much more complicated, and the slope is therefore not a constant. The usual practice of TV engineers is to copy an accepted measure in photography, where the Hurter-Driffield or D-log E curve is characterized by the slope gamma in the inflection point of the ogive. That is the reason why in practice this slope, or rather the slope of the approximately linear part of the luminance-reproduction curve of the imaging chain on a double log basis, is often used as the characteristic parameter for brightness contrast. However, in TV images, especially when they involve scanning slides or motion picture films, the luminance-reproduction curve can be quite capricious due to a combination of factors such as the tone reproduction of the film, the monitor characteristics and will often be distorted by the effect of reflected ambient light. An example of measured tone reproduction of slides is shown in Figure l. The slope obtained by simple linear regression of the middle part of the ogive need not to be an entirely adequate measure, as will be explained below.

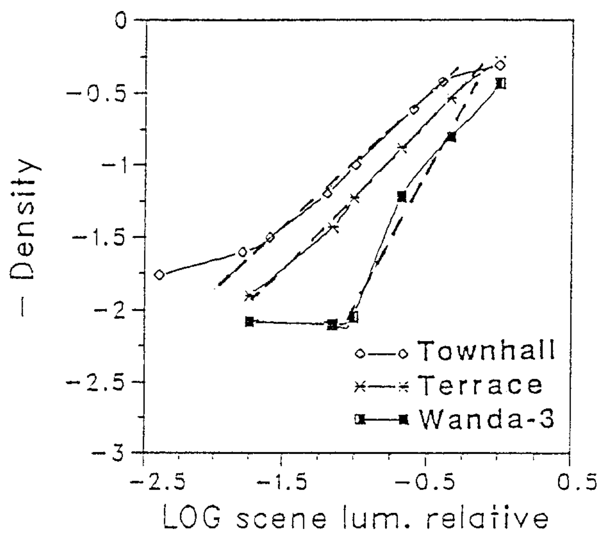

Figure 1. Example of the tone reproduction of slides used in the experiments. The different curves correspond to different scenes, which underwent different photographic processing.

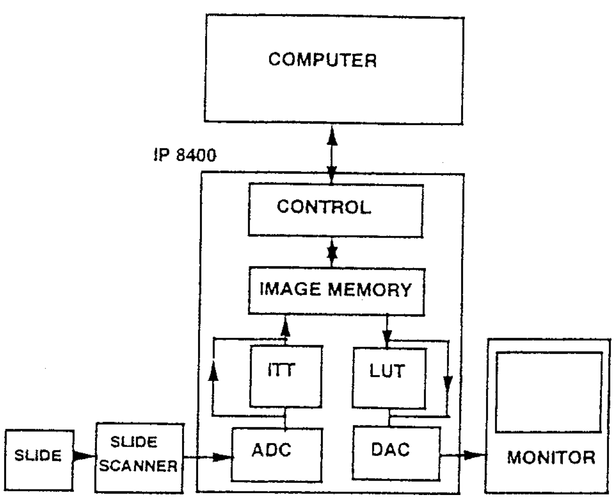

Figure 2. Diagram of the image-processing set-up.

At this point it seems appropriate to point out that in principle there is a difference between an adequate measure of brightness contrast in relation to the potential of the display to transfer brightness contrast for any possible scene on the one hand and an adequate measure of the actual brightness contrast of a particular scene on the other. The latter is needed to relate brightness contrast and image quality, which is the major point of interest in this paper.

Roufs and Goossens,[3] used still black-and-white images to show that the relation between perceptual quality and the parameter gamma is mostly determined by perceived brightness contrast. The results confirmed those of Bartleson and Breneman[5] in the sense that the optimal value of gamma was found to be larger than 1 for all test scenes. Bartleson and Breneman argued that this is caused by a change in the brightness pattern if the scene is looked at in a different surrounding than the original one, especially if this surround is relatively dark, as is the case with slides and TV. However, Bartleson and Breneman did not report different optimal values of gamma for different scenes. Although the results we obtained are quite consistent between subjects, the optimal values of gamma of some of the test scenes were found to be significantly different. We will first show that new experiments with still scenes both black-and-white and identical coloured gave results which are consistent with the earlier findings. Moreover, a refined analysis of the luminance reproduction in combination with the grey-level distribution of the various test scenes shows that an effective gamma can be defined. Using this effective gamma the quality versus gamma plots of all scenes coincide. This is basically a result of the coincidence of the brightness contrast versus gamma curves for different pictures.

This leaves us with the question of how general these findings are in view of the effect of possible differences in grey-level distributions over area elements of different arbitrary scenes. This matter is tackled by scaling brightness contrast as a function of gamma using gaussian blobs of randomly distributed sizes, grey-level distributions and polarities instead of natural scenes.

These stimuli have the advantage that the ratio of the area of dark and light spots can be changed easily without introducing much other problems, as will be explained below.

Apparatus and Methods

All experiments were performed with the set-up in Figure 2. Slides of natural scenes were scanned by a flying spot scanner. The signals passed a compressing gamma-correction network, which transformed the amplitude of the signals according to a power function, the power being 0.4. In a Gould deAnza IP 8400 processor, connected to a micro-VAX, the signal was converted into 8-bit digital grey levels (ADC) and put into a memory. The digital grey level information was processed according to a power function the exponent of which could be controlled. Look up tables (LUTs) were used to correct the data for the deviations of the monitor characteristic from the power function in the working range. An iterative algorithm ensured that, when the exponent of the transfer function was changed, the mean luminance was kept constant by changing the peak luminance. The reason was to avoid interaction caused by variations in mean luminance, since mean brightness was found to be another dimension of perceptual image quality.[6] The algorithm also prevented the peak in the distribution being shifted to gray levels which are larger than the 8-bit range allows. The consequence of these iterative procedures is that the final mean luminance of all processed versions of one scene, although constant, cannot be fixed a priori. After D A conversion the signals were transformed into images by means of high-quality monitors (Conrac 7211 for experiment 1; Barco CTVM or Philips LDH for experiment 2; and Barco CCID 735 for experiment 3). Their luminance-grey level relations were corrected by LUTs to obtain an overall relationship that is a power function with an exponent equal to 2.5. The luminance-reproduction function of the complete chain could be described by a power relation. The power can be controlled by the IP 8400, if the Hurter-Driffield curve of the slide also matched a power function. Unfortunately, the slides showed considerable deviations from such a powerfunction behaviour. They were calibrated by placing a board with known diffusely reflecting grey steps in the scene during one of the two identical shots, which were taken in immediate succession. Of the monitor surface area of 30 × 40 cm, only 28 × 28 cm containing 512 × 512 pixels was used. Unless stated otherwise, the distance of the observer was 2.1 m (7H).

The measurements of experiments 1 and 2 were performed in a dark room with faint illumination of the wall behind the monitor (4 cd m^{-2}). The mean luminance varied for the different test scenes. On the average it was about 25 cd.m^{-2}. If not stated otherwise, only the black-and-white versions of the images were used. The pictures were outdoor scenes and portraits of a female model.

Ratings of subjects were, if not stated otherwise, based on a 10-point scale of numerical categories. Since we never found deviations from equality of distances between categories under these conditions, the raw data are given here. Subjects did not rate the strength of both psychological attributes (overall perceptual image quality and global brightness contrast) in one session. The stimuli were presented in random order for 5 seconds at a time, after which (in most cases) a neutral grey of the same luminance as the test-scene was presented for 2-seconds. Every test stimulus was repeated either 6 or 8 times.

At the beginning of each session the subjects were presented with a trial series which contained all extremes and enabled them to establish their internal scale sensitivity. All subjects had normal visual acuity. They were either students or members of the staff of the institute. Half of them had experience of scaling experiments based on visual stimuli.

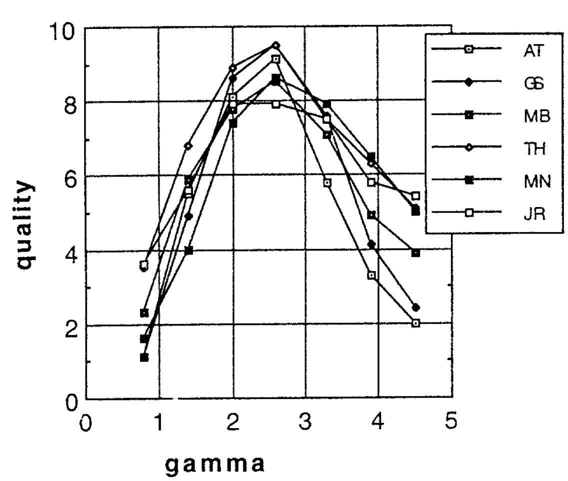

Figure 3. Quality ratings as a function of gamma of 5 subjects for one of the scenes.

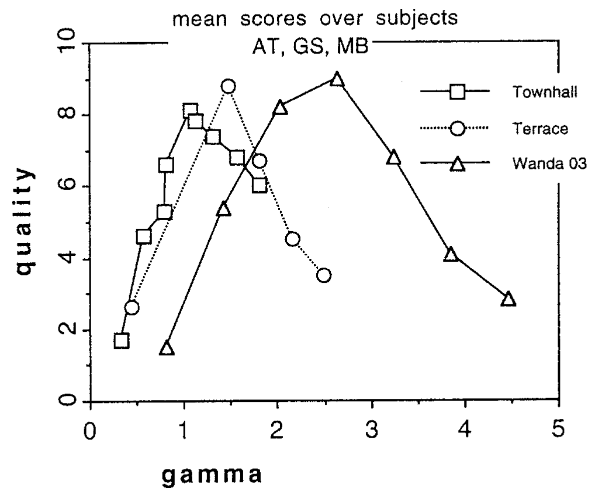

Figure 4. The mean quality scores as a function of gamma over 3 subjects and for 3 scenes. Wanda 03 is the portrait of a female model, the others are outdoor scenes. mean scores over subjects

Results

Experiment 1:

Scaled Perceptual Quality and Brightness Contrast versus Gamma. Figure 3 shows examples of results of quality judgements as a function of gamma, performed by different subjects for one scene. The mean quality scores of 3 subjects are plotted in Figure 4 for 3 scenes. The curves show a clear optimum, which is not significantly different between subjects, but which seems to differ significantly from one scene to another. In retrospect, the subjects said that their quality judgement probably had a lot to do with the naturalness of the pictures displayed. The effect of colour on the quality judgements is studied by transforming the Y signal according to the power function and adding the original UV colour signals (PAL coding). The results are shown in Figure 5. The positions of the peaks should be compared with those of Figure 4. Colour obviously does not have a large effect on the judgements.

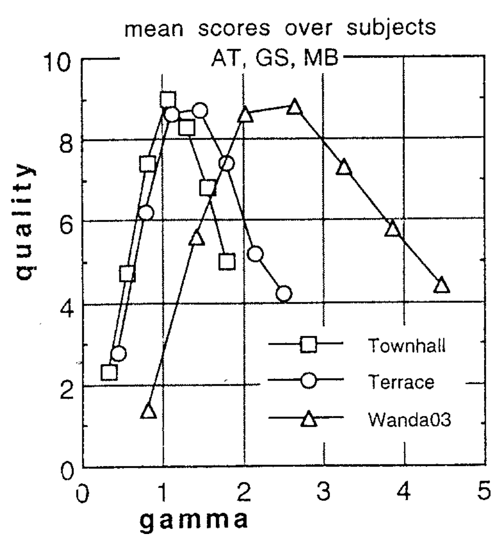

Figure 5. Perceptual quality ratings by the same subjects as a function of gamma of the same scenes but this time coloured.

Figure 6 shows brightness contrast judgements of the same black-and-white pictures as in Figure 4. Those ratings were made by the same subjects, though, in a separate session. As expected they increase monotonically with the parameter gamma. Their dependence on gamma is different for different scenes. If rated perceptual image quality is plotted against rated brightness contrast, however, all curves come together, as demonstrated in Figure 7. The almost complete coincidence of the curves suggests that global brightness contrast is the dominating sensorial dimension under the present conditions.

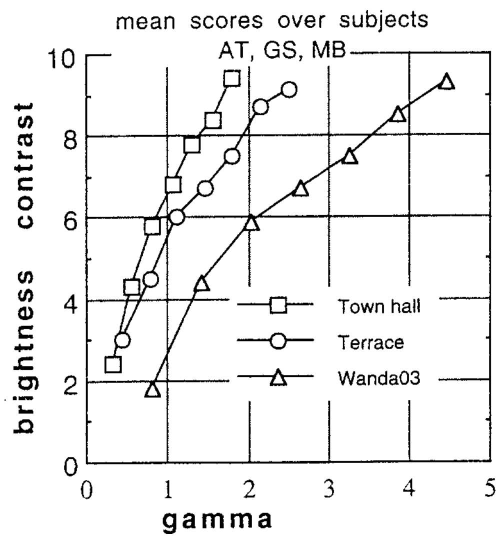

Figure 6. The ratings of global brightness contrast as a function of gamma of the black-and-white pictures of Fig 4.

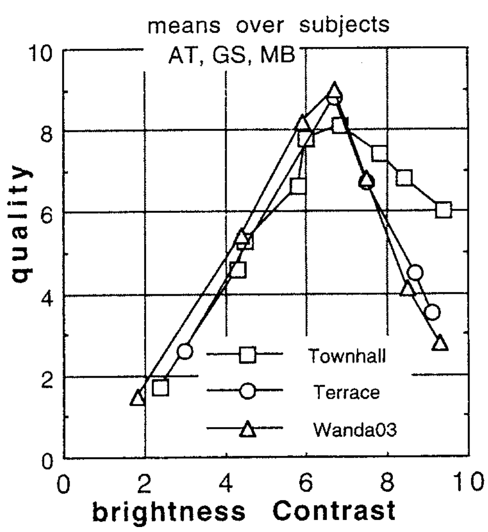

Figure 7. Scaled perceptual quality as a function of scaled global brightness contrast for 3 scenes.

Experiment 2:

The Effect of Observer Distance on Global Brightness Contrast. In view of our interest in the generality of the above observations, we changed the observer-screen distance. The observers were positioned 1.2, 2.1, 4.5 and 9.0 m from the screen. Another reason to perform these experiments is that on the one hand, recent models on the brightness-luminance relation of luminance patterns[7] have the property of invariancy for distance, although the mean global brightness is fairly strong influenced by the far peripheral visual field.[8] On the other hand in view of the theory of Bartleson and Breneman[9] the dark surround would cause the change in brightness pattern, which would be the reason for our preference for gamma greater than one. This makes it interesting to see, whether global brightness contrast is distance-invariant and therefore also the optimal value of gamma.

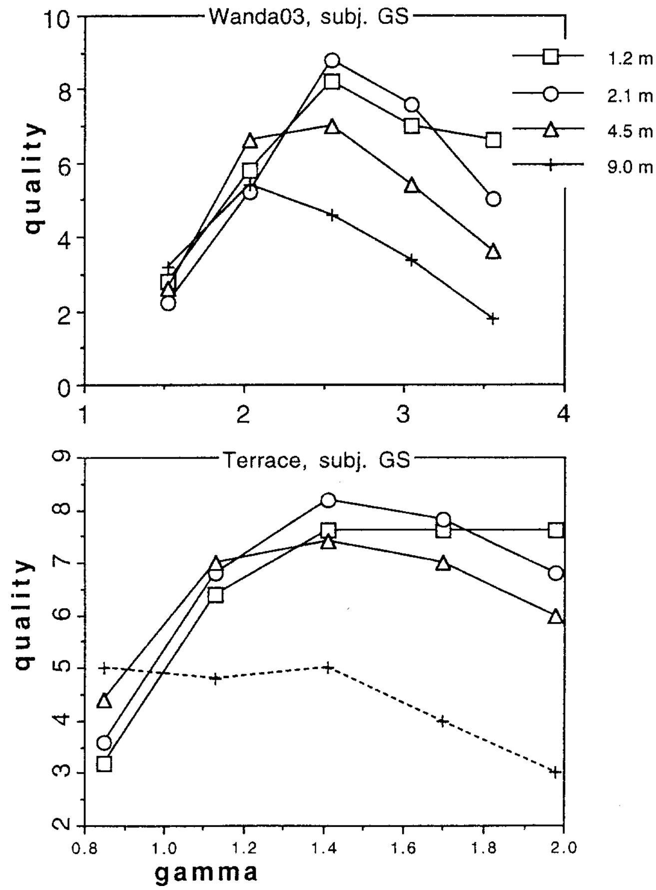

Figure 8. Examples of scaled perceptual quality of one subject at different observer distances, having one constant reference image available at 2.1 m and containing the same scene with the same gamma as the test stimulus.

In the present experiments we wanted to avoid any reestablishment of the scale sensitivity at the various distances in order to be able to check the findings with earlier data. Therefore, the subjects had in all cases a reference image with an identical scene at 2.1 m distance. An example of the results of one subject and two scenes is given in Figure 8. In view of the spread, we can conclude there is no significant shift of the optimal gamma values with distance, indicating no variations with distance for the value of the optimum within the experimental spread. The quality ratings decrease at larger distances, in agreement with earlier literature findings.[10] In Figure 9 this is demonstrated by plotting the mean quality scores of 3 subjects as a function of distance. It would have been quite interesting to be able to work with very large images since one would expect the optimal value of gamma to be 1 again. Unfortunately, we cannot simulate this condition by using very short distances, since image degradations such as the line or the dot structure then become visible. Since in the present conditions changes in perceptual quality are linked with changes in brightness contrast, the invariance of the optimal gamma values with respect to distance would fit in with the scale-invariance of the brightness pattern mentioned above. The expected decrease in quality judgements for short distances (i.e. where the TV lines or dots become visible), is still limited at a distance of 1.2 m.

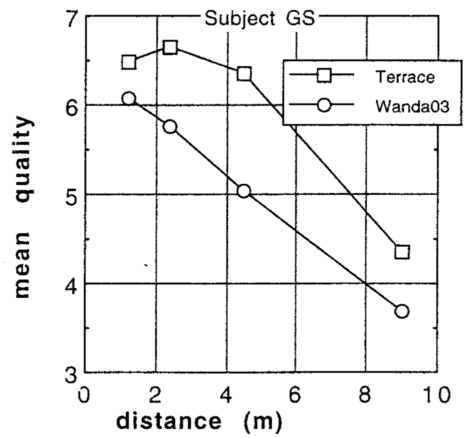

Figure 9. The mean rated quality of 3 subjects as a function of observer distance.

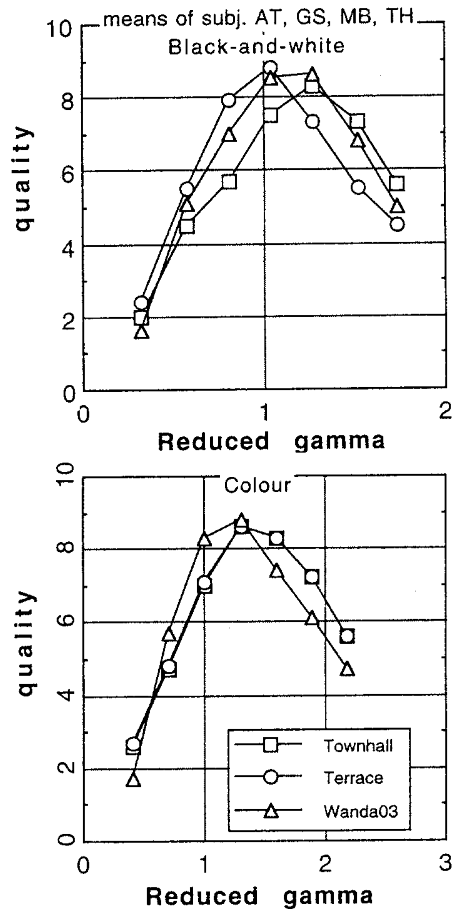

Figure 10. Scaled perceptual quality of Figs 4 & 5 as a function of gamma of the slide-scanner-monitor chain.

Modification of Gamma:

A Reconsideration of Experiments 1 and 2. A first indication of what might be behind the different optimal values of gamma for different scenes was found in the observation that the gamma of the slide scanner - monitor chain was about 1.2 in all cases. Fig 10 demonstrates this.

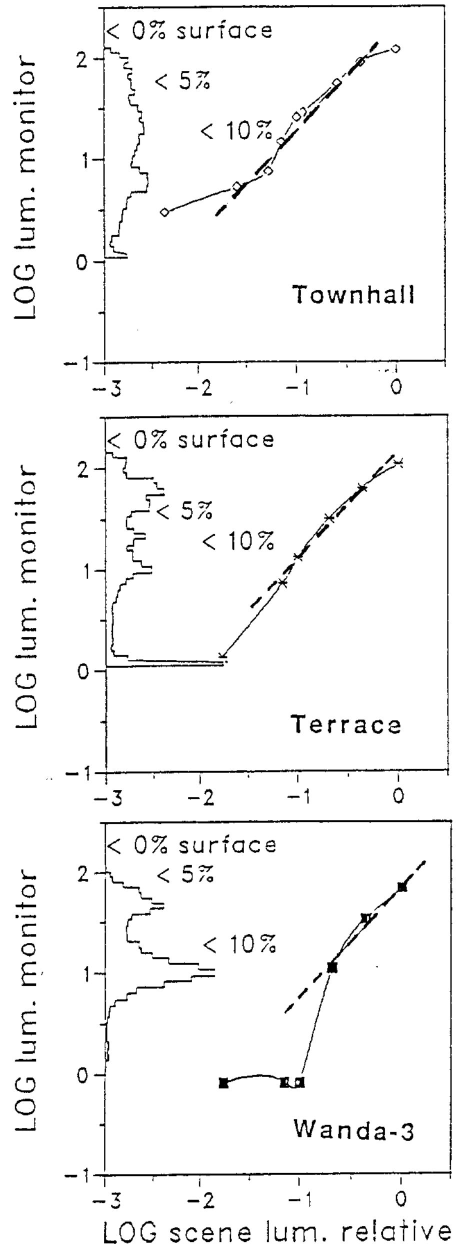

Figure 11. Grey-level distribution and luminance reproduction function of the imaging chain.

A refined analysis of the grey-level distribution and the known properties of slide film brought the second clue. Figure 11 shows the luminance-level distribution of some typical scenes plotted together with the luminance-reproduction curves. It is clear that the local slope of this function at positions where most grey levels of the scene occur can differ quite considerably from the slope determined by a simple linear approximation of this function through the measured levels of the calibration patches. Using a kind of effective gamma by determining the slope of that part of the luminance-reproduction function that is most frequently occupied (weighted linear regression) does indeed bring together the rated global brightness contrast (Figure 12) and perceptual quality curves (Figure 13) of the different scenes as a function of the effective gamma. The optimal value for the effective gamma is about 1.2-1.3. For a more definite value more tests would be needed. Anyway, this value is very near what Bartleson and Breneman[9] found for TV in 1967. One can only admire their achievements, since they had to work with so much more primitive means.

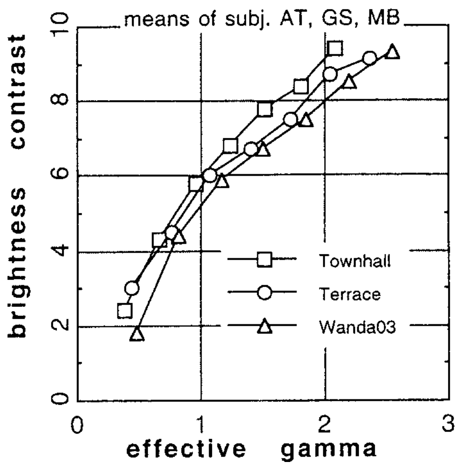

Figure 12. Mean scaled global brightness contrast of 3 subjects as a function of the effective gamma for the three different scenes.

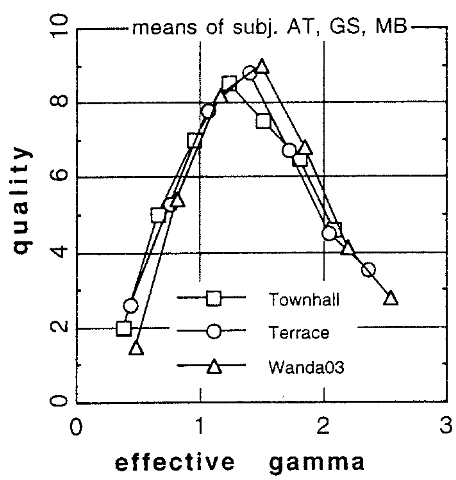

Figure 13. Mean scaled perceptual quality of 3 subjects as a function of the effective gamma for the 3 different scenes.

Experiment 3:

The Effect of Luminance Distribution on Global Brightness Contrast. The importance of global contrast again raises the question of the significance of the distribution of luminance over area for the judgement of brightness contrast. To investigate this, we used test stimuli which enabled us to change this distribution. In order to avoid artefacts by memory (*i.e.* caused by the subjects' inclination to repeat the rating of a scene with certain parameter values which they remember having seen before), we chose random stimulus. We also wanted to avoid cues in other sensorial dimensions like sharpness. Blobs with a gaussian luminance profile with different diameters, different modulation depths and both polarities were found to be a reasonable solution with respect to these specifications. Examples of histograms of luminance-level distributions are shown in Figure 14. Figure 15 demonstrates the appearance of the stimuli used in these experiments for 4 different distributions.

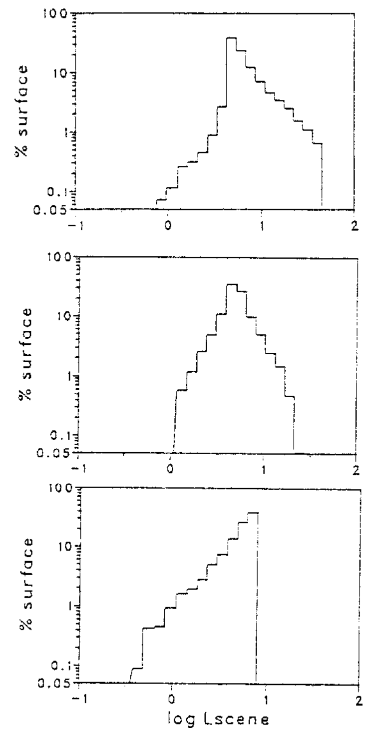

Figure 14. Examples of grey-level distributions of used stimuli from top to bottom more: white than dark blobs, symmetrically distributed white and dark blobs, more black than white blobs. Different luminance ranges were used, see text.

Global brightness contrast was scaled with numerical categories 0-8, half-points being allowed. Data of 6 subjects were averaged over their Z-transforms, which are a result of the calculated sensorial scale value according to Thurstone's model D. We chose $C_{LR} = \log (L_{max}/L_{min})$ as the luminance contrast parameter. This measure should be proportional to gamma if the luminance range is transferred according to a power function. Figure 16 shows the results for a symmetrical distribution coded eqno (as many bright blobs as dark ones), for a distribution with bright dots in the majority, (coded whno) and finally a distribution with a majority of dark blobs (coded blno). Although there is some effect of the skewness of the distribution, it is small compared with that of the luminance range, implying that a brightness contrast impression is not very sensitive to the nature of the scene. This may be appreciated if one compares the magnitude of the effect due to these large differences in skewness, being less than 1 point out of 10, and estimate the resulting shift of the quality function of a few tenth in Figure 5 from the shift in Figure 6. It confirms the insensitivity of scaled brightness contrast to the nature of the scene if effective gamma was used as a parameter. (It does not say that contrast would not change considerably with extreme distributions, as one would intuitively expect.) Finally, brightness contrast averaged over subjects of the eqno distribution (Figure 17) can be described empirically by:

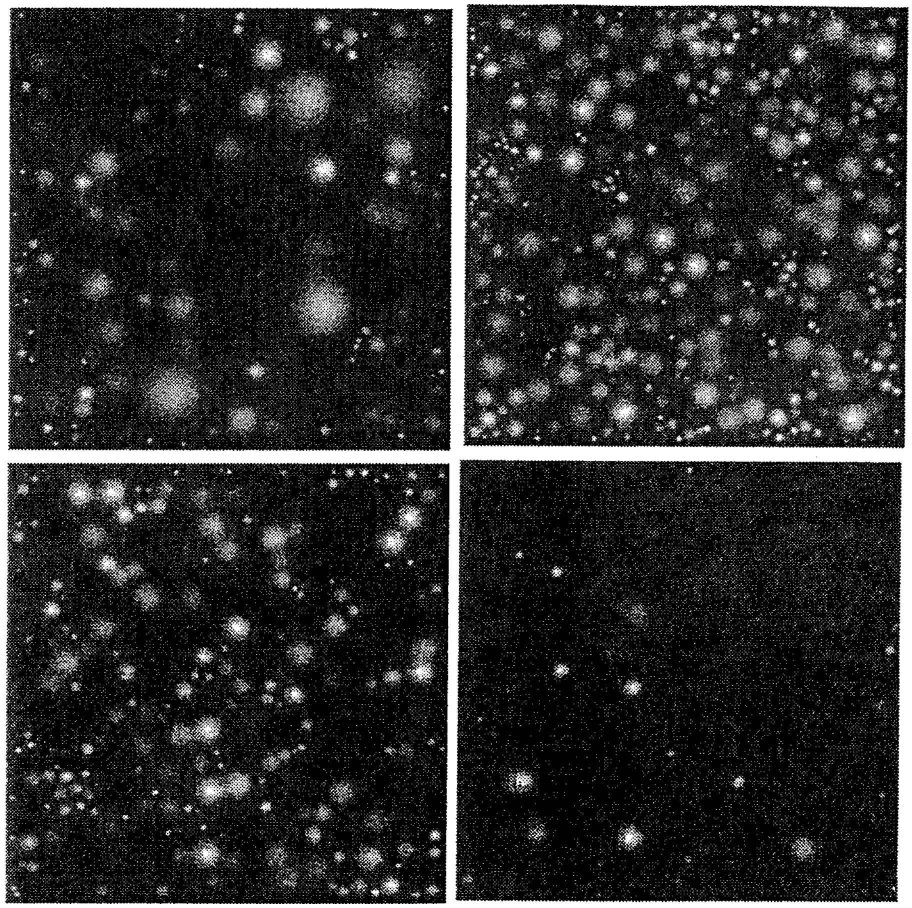

Figure 15. Examples of the gaussian blob test scenes. Upper left: symmetrically distributed large dark and light blobs (eqla). Upper right: light- dominated normal-sized blobs (whno). Lower left: symmetrically distributed, normal-sized blobs (eqno). Lower right: dark-dominated distribution of normal-sized blobs (bnno).

$$C_b = a\, C^{\alpha}_{LR} + b,$$

α being about 0.5 (C_b = brightness contrast, C_{LR} = luminance contrast defined by the ratio of the maximum to minimum luminance in the scene. This is consistent with literature data of sinusoidal gratings.[11]

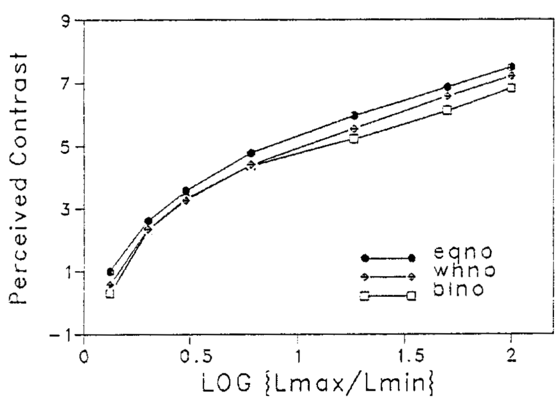

Figure 16. Scaled global brightness contrast averaged over the Z-transforms of the subjects, obtained from the Thurstonian model D, as a function of the luminance ratio range. Three distributions with different skewness are used.

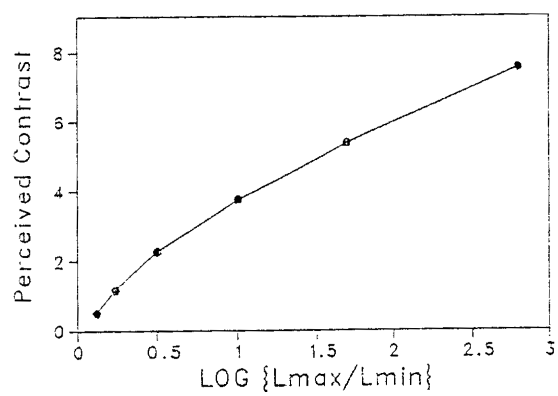

Figure 17. Scaled global contrast of a symmetric distribution as a function of the luminance ratio range. The line is fitted according to the formula in the text, the exponent being 0.5.

Conclusions

- Repeated experiments with still black-and white TV images on the relation between perceptual quality and gamma confirmed earlier results. There is a clear optimum in scaled perceptual quality and consistency between subjects. However, the raw optimal gamma values for different scenes differ considerably.

- If scaled perceptual quality is plotted against scaled global brightness contrast, all curves coincide. This implies that global brightness contrast is the dominant factor for quality under the present conditions.
- The optimal value of gamma is not influenced by colour and independent of observer distance, within the limits of the experimental accuracy.
- There are strong indications that the differences between scenes is due to the fact that gamma, derived from a linear approximation of the middle part of the luminance reproduction function is not quite adequate. If gamma is determined in such a way that it reflects the slope of the luminance-reproduction curve in parts which are relevant to the luminance distribution, the curves for the different scenes also coincide. The optimal corrected gamma value was found to be 1.2-1.3.
- The lack of sensitivity of global brightness contrast to the luminance distribution over area emphasizes the usability of the measurement of the strength of global brightness contrast for any natural scene.
- If we neglect the effect of area, the log of the ratio of maximum to minimum luminances seems to be an adequate physical measure for global brightness contrast of a complex scene, a measure which is not unfamiliar to TV engineers.

References

1. J.W. Alnatt, N. Gleis, F. Kretz, A. Sciarapa and E. VanderZee, "Definition and validation of methods of subjective assessment of visual telephone picture quality,". *CSE2T Raporti technici*, Vol. **XI**, pp.59-65, 1983.
2. J.A.J. Roufs, "Perceptual image quality: concept and measurement," *Philips J. Res.*, Vol. **47**, pp. 35-62, 1992.
3. J.A.J. Roufs and A.M.J. Goossens, "The effect of gamma on perceived image quality," *Proc. 1988 Intl. Display Conf.* San Diego, CA *IDRC Digest 1988*, IEEE, NY, pp. 27-31, 1988.
4. J.A.J. Roufs, "Brightness contrast and sharpness, interactive factors in perceptual image quality," In: *Human Vision, Visual Processing and Digital Display, Proc. SPIE*, Vol. **1077**, pp. 66-72, 1989.
5. C.J. Bartleson and E.J. Breneman, "Brightness reproduction in the photographic process," *Photogr. Sci. Eng.*, Vol. **11**, pp. 254-262, 1969.
6. M.H.W.A. Boesten and E. VanderZee, "Psychophysical versus psychometric methods in image quality measurements," *IPO Annual Progress Rep.*, Vol. **16**, pp. 67-71, 1981.
7. F.J.J. Blommaert and J.B.O.S. Martens, "An object-oriented model for brightness perception," *Spatial Vision*, Vol. **5**, pp. 15-41, 1990.
8. G. Schouten, Luminance brightness mapping: the missing decades, *Thesis* Eindhoven University of Technology, 1993.
9. C.J. Bartleson and E.J. Breneman, "Brightness perception in complex fields," *J. Opt. Soc. Am.*, Vol. **57**, pp. 953-957, 1967.
10. J.H.D.M. Westerink and J.A.J. Roufs, "Subjective image quality as a function of viewing distance, resolution and picture size," *SMPTE J.*, Vol. **98**, pp. 113-119, 1989.
11. M.W. Cannon, "A study of stimulus range effects in free modulus magnitude estimation of contrast," *Vision Research*, Vol. **9**, pp. 1045-1052, 1984.

published previously in SPIE, Vol. 2179, page 80