# Subjective picture quality as a function of viewing distance, picture size and resolution

JP/ML
No. 202839

88/01
11.01.1988

Rapport no. 627

# Subjective picture quality as a function of viewing distance, picture size and resolution

J.H.D.M. Westerink

This report is an English version of IPO-report no. 586.

**Summary**

This report describes two experiments concerning the subjective quality of complex scenes. Use was made of slide projections and these were varied in respect of resolution, viewing distance and picture size. The subjective quality was judged by a group of twenty subjects by means of categorical scaling.

The results of the experiments show that the absolute resolution expressed in periods per degree and the picture angle spanned by the display each influence the quality independently. The subjective quality increases with the resolution, but saturates at a resolution (6 dB cut-off frequency) of approximately 25 periods per degree. There is also a linear relationship between the subjective quality and the logarithm of the picture angle.

In the discussion these results are compared with a number of experiments known from the literature. The results are also translated to a number of practical situations, for example the consequences for the use of High Definition TV.

Contents Page

## 1 Introduction

This report describes a number of experiments relating to picture quality as a function of several physical parameters. Let us therefore first consider the concept of picture quality.

The concept of picture quality is rooted in two different worlds. The **picture** is to be described on the basis of its physical parameters, such as its size, luminance, resolution, spectral scene content and artefacts such as noise and flicker. Together these parameters define a multidimensional physical space.

The **quality** of the picture is however subjective, that is to say dependent on the person watching it. It must therefore be described in psychological terms, known as percepts or sensations. Brightness, sharpness, subjective contrast and perceived size are examples of this. In their turn the percepts also define a multidimensional (psychological) space. Picture quality is thus to be seen as a function with as its domain the physical space and as its range the pyschological.

The general aim of the 'Picture Quality' research project is to arrive at a characterisation of the (subjective) quality as a function of various global physical parameters. Artefacts, which usually have a local character in the picture, are for the time being left out of consideration. In previous experiments at the IPO the influence of luminance, picture size and viewing distance on the quality was studied by van der Zee and Boesten (13), (14). For complex scenes with a high resolution they found an increase in the subjective quality with increasing luminance and with increasing picture size. The experiments in the present report are a direct continuation of their work and relate to the influences of resolution, picture format and viewing distance.

Although these experiments are simple in terms of test configuration, they are not as yet reported in the literature. The reason for this may be that the technicians are essentially interested in the physical description of the picture, while 'quality' is too vague a concept for psychologists. Experiments in which some, but not all the present parameters are varied, are indeed known.

Van der Zee and Boesten (14), for example, only varied viewing distance a and picture width b in the case of complex scenes with a high resolution. According to their results the quality increases with the logarithm of $b^2/a$ ,which they explain as a consequence of the known 'size-constancy effect'. Measurements by Hatada (5) point in the same direction.

The influence of resolution on picture quality has, as appears from the literature, been studied repeatedly. Reference can be made to studies by Snyder (9), Higgins (6), Task (10) and many others (see literature study by Westerink (12)). They describe that the picture quality first rises with increasing resolution and subsequently saturates as a consequence of the spatial sensitivity of the visual system. What these studies have in common however is that as a rule they only vary the resolution and not the viewing distance or the picture width, so that a description of the interaction of these parameters is not possible.

Such a limitation also applies to the data of Jesty (7), who determined the optimal viewing distance dependent upon picture width and resolution. He finds that the quotient of this optimal viewing distance and the picture width is constant and dependent on the (relative) resolution of the picture. How the quality progresses beyond this optimal viewing distance cannot however be deduced from these experiments.

In conclusion it can be stated that the experiments described in the literature constitute various cross-sections of the experiments of this report. This study describes the quality in the 'complete' subspace which is defined by the three parameters of resolution, picture width and viewing distance. On the basis of the results it can therefore be established whether, and in what way, the literature data are to be interrelated.

# 2 The experiments

## 2.1 Method

Pictures are used in different environments and the concept of picture quality can accordingly be dependent on that environment. Hence the picture quality of military aerial photographs, of medical X-ray photographs or of text displays are essentially characterised by the possibility for, or the ease of, extracting information. This is also referred to as picture quality in a task-oriented environment. In such environments it is possible to have subjects perform a task which is concerned with the information in the pictures. A measure of performance, such as the average percentage correct score or the reaction time can then serve as a measure for the subjective picture quality.

However, it is on the contrary the picture quality in non-task-oriented environments which interests us. This means that it is not possible to have subjects perform a task without running the risk that the attitude of the subject towards the picture will change. For this reason a performance measure is not one of the possibilities for expressing subjective quality and we must confine ourselves to the **judgment** of the subject.

In the experiments described the judgment of the subject is recorded by means of the categorical scaling method. A subject is presented with a stimulus and he is asked to assess its quality by placing the stimulus in one of the available categories of the categorical scale. In these experiments a 100-point categorical scale is used, which runs from 0.1 to 10.0, and thus corresponds with the Dutch system of report marks.

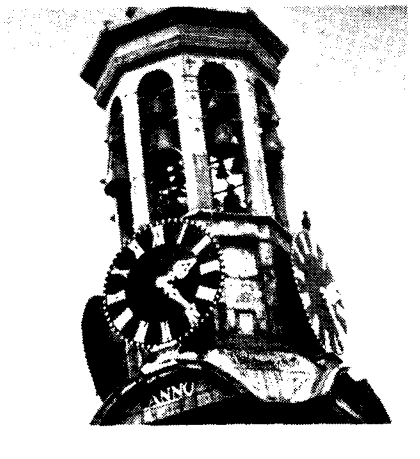
The categorical scale values collected for all subjects and all stimuli are termed the rough data. They describe the stimuli (and subjects) on the basis of a linear objective scale: the difference between the categorical scale values 7.0 and 8.0 is the same as that between 2.0 and 3.0. It is not necessarily the case however that the difference in quality between a 7.0 and an 8.0 is actually perceived as being as great as the difference in quality between a 2.0 and a 3.0. In other words: the categorical scale is not necessarily linear in psychological terms, whereas it is precisely such a psychologically linear scale which interests us. This can however be constructed from the rough data on the basis of

a)

b)


c)

d)



e)

Figure 1: Scenes used
The above scenes are black-and-white reproductions of the colour slides used in the experiment.

a) Ropes
b) Terrace
c) Graffiti wall
d) Thielke
e) Tower of the Mint

All scenes are used in EXP1; scenes a to d are used in EXP2.

the spread therein according to a model by Thurstone (Law of Categorical Judgment, see for example Torgerson (11)). The rough data are thereby converted by means of a non-linear transformation to a psychologically linear scale. These scale transformations are discussed in further detail in chapter 3.

## 2.2 Test configuration

Two experiments were carried out; they differ essentially in the parameters over which the stimuli were varied:

- EXP1: The variables are picture width and resolution. The picture height varies along with the picture width, so that the scene is always square. The resolution $f_{abs}$ is calculated from the 6 dB cut-off frequency of the modulation transfer function of the display equipment and expressed in periods per degree ( ~ /o; further details of the measures of resolution applied are given in paragraph 2.3). The viewing distance was 2.9 m. Table 1 indicates what combinations of resolution and picture width are used, with every picture width/resolution combination being shown for five different scenes (see figure 1), which in total produces 140 different stimuli. Twenty subjects, with a vision of at least 1.0, and all students and employees of the IPO, participated in the experiment.

- EXP2: The variables are picture width, resolution and viewing distance. Table 2 indicates what combinations of resolution, picture width and viewing distance are used. Due to an anticipated loss of concentration on the part of the subjects, the series of stimuli cannot be too long. For this reason a choice was made from all possible combinations of variables and thus the experimental configuration is not completely crossed. All combinations from table 2 are shown for four different scenes (see figure 1), which resulted in a total of 112 presentations. Twenty subjects, again with a vision of at least 1.0, and all employees or students at the IPO, participated in EXP2. Four of them also participated in EXP1.

Table 1: Stimulus data in EXP1

On the vertical scale are the spatial resolutions $f_{abs}$, expressed in ~/o. On the horizontal axis are the picture widths b, expressed in centimetres. The cells of the table indicate how often this combination for a certain scene was presented to a subject. The viewing distance was in this experiment 2.9 m.

| $f_{abs}$/b | 92 | 72 | 48 | 24 |
|---|---|---|---|---|
| 38 | 1 | 1 | 1 | 1 |
| 26 | 1 | 1 | 1 | 1 |
| 19 | 1 | 1 | 1 | 1 |
| 14.0 | 1 | 1 | 1 | 1 |
| 8.2 | 1 | 1 | 1 | 1 |
| 3.9 | 1 | 1 | 1 | 1 |
| 2.7 | 1 | 1 | 1 | 1 |

Table 2: Stimulus data in EXP2

In each of the tables the spatial resolutions $f_{abs}$, expressed in ~/o, are indicated on the vertical axis. On the horizontal axis the picture widths b, expressed in centimetres are indicated. The cells of the table indicate how often this combination for a certain scene was presented to the subject. For each viewing distance there is a separate table.

a)viewing distance 2.9 m

| $f_{abs}$/b | 92 | 72 | 48 | 24 |
|---|---|---|---|---|
| 43 | 1 | 1 | 1 | 1 |
| 24 | | | | |
| 8.7 | | | | 1 |
| 2.6 | 1 | | | |

b)viewing distance 3.9 m

| $f_{abs}$/b | 92 | 63 | 31 |
|---|---|---|---|
| 28 | 2 | 2 | 2 |

c)viewing distance

| $f_{abs}$/b | 92 | 72 | 48 | 24 |
|---|---|---|---|---|
| 59 | 1 | 1 | 1 | 1 |
| 33 | 1 | | 1 | |
| 23 | 1 | | 1 | |
| 8.7 | 1 | 1 | 1 | 1 |
| 2.7 | 1 | | 1 | |

The stimuli were set for the two experiments in a 'quasi' random sequence. It was ensured that stimuli with a high quality were evenly distributed in the series. Nor was the same scene ever shown twice in succession. Nor was a certain combination of the variables picture width, resolution and viewing distance ever followed twice by the same other combination. For each subject the sequence of the stimuli was the same, but the starting point in this series was chosen arbitrarily. Thus one subject first did the last part of another's and then his beginning. As a result, averaged over the subjects, all stimuli occur equally often at the beginning as at the end of the series.

Prior to the series of stimuli the subjects were presented with a number of 'trial' stimuli. These trial stimuli cover the complete range of the subsequent 'real' stimuli in terms of quality and are also termed anchor points. They are designed to give the subject an impression of the stimuli to be expected, so that he can adapt his use of the categorical scale accordingly. The judgments of the trial stimuli were not included in the processing of the results.

The details were related to the subjects at the beginning of the experiment in the form of a written text (see appendix A).

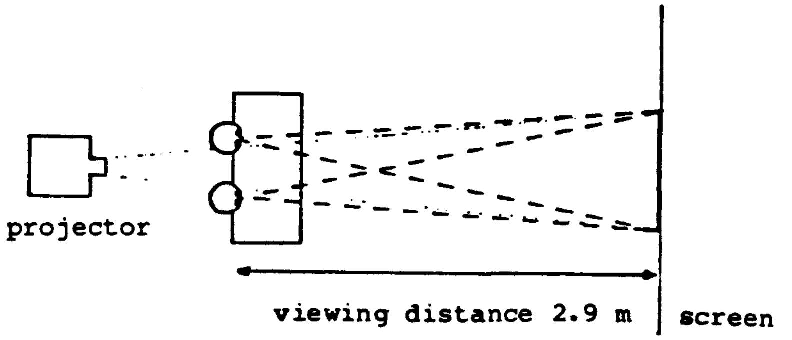
### 2.3 Implementation

The stimuli were constituted by projections of slides with complex scenes onto a diffuse projection screen. For the studying of (TV) picture quality this has the disadvantage that the pictures do not move. In the case of still pictures the emotional involvement in the pictures is perhaps slightly less and the sensitivity to local details possibly slightly greater. The still picture is thus to be considered as a 'worst-case' approximation of the moving picture.

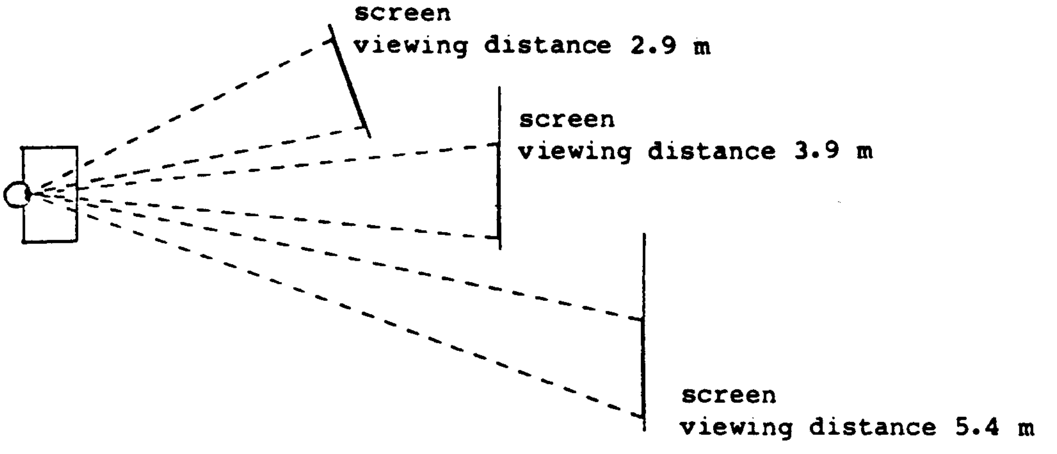
A great advantage in working with slide projections is the fact that it is possible to vary a stimulus parameter over a wide range, without any other parameter varying with it. The picture can, for example, be enlarged without the noise increasing.

For projection Kodak Carousel S-RA 2000 projectors were used, equipped with a Leitz 150 mm lens. The light intensity of the projector lamps was regulated by separate power supplies with a view to current stabilisation.

The variations of the different parameters were brought about as follows.




a)




b)

Figure 2: Test configuration
a) Test configuration for EXP1
b) Test configuration for EXP2.

- Viewing distance: Use was made of a number of projection screens, which were positioned at various viewing distances from the subject. Each projection screen had its 'own' projector. In EXP2 an identical object (in casu a chair) was positioned in front of each projection screen in order to give the subject a cue for the viewing distance. A sketch of the configurations is provided in figure 2.

- Picture width: The picture width was varied by copying the originals of the scenes onto various formats. In this way it was possible to change the picture width without moving the projector.

  The original scenes were recorded on 6x6 $cm^2$ large master positives. These were reduced on high-resolution slide duplicating film to the formats 4x4, 3x3, 2x2, and 1x1 $cm^2$. The final resolution of the slide copies thereby remained greater than the resolving power of the projector lens.

- Resolution: The resolution was varied by defocusing the lens of the projector. For this purpose the lens was mounted on an x-y table, which was positioned through the combination of a stepping motor, an interface and an Apple IIe.

  For different lens positions the stepwise response of the projection system was measured. From this the modulation transfer function (MTF) of the system was determined via Fourier transformation. The 6 dB cut-off frequency $f_{6dB}$ of this MTF was taken as a measure for the resolution. It was expressed in periods per metre-on-the-screen ( ~ /m), and at a known viewing distance a can be converted into periods per degree ( ~/o), which produces the (absolute) resolution: $f_{abs} = 2\ /360.f_{6dB}.a$. It must be emphasised that this absolute resolution is independent of the picture width of the slide.

  The variation of the $6_{dB}$ cut-off frequency applied in this way proved to be repoducible and more or less uniform over the entire plane of projection (for more details see appendix B).

The stimuli had a presentation time of 16 s (15 s in EXP2). The inter-stimulus time was 19 s (5 s in EXP2), in which a uniform white plane was projected with a luminance of 30 cd/m$^2$ (25 cd/$^2$ in EXP2). This plane served to diminish any after-images left with the subject, while he nevertheless remained adapted to the stimulus luminance. The average luminance of the scenes was around 30 to 40 cd/m$^2$; the open-gate luminance of the projectors was of course higher: 320 cd/m$^2$ (in EXP2 280 cd/m$^2$). The projectors supplied the sole lighting in the room, as a result of which the ambient luminance had a value of approximately 5 cd/m$^2$.

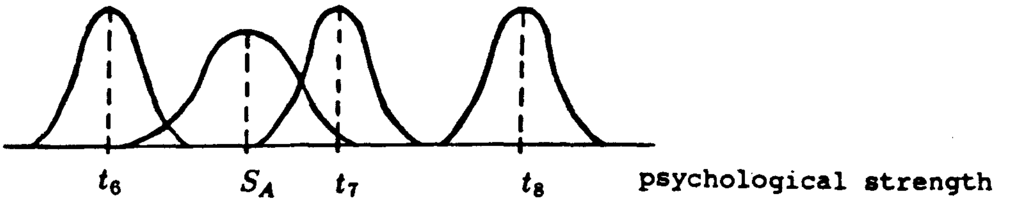



Figure 3: The psychological continuum
Three different average category limits $t_6$, $t_7$ and $t_8$ are indicated along with the normal distributions to which they belong. The average strength of the impression of a certain stimulus $S_A$ is also indicated, along with the distribution to which this belongs.

## 3 Categorical scales and subjective scales

### 3.1 Thurstone's Law of Categorical Judgment

Thurstone's 'law of categorical judgment' describes the judgments of a subject when the latter has to describe a certain impression (for example the impression of quality) on a categorical scale. Thurstone assumes that the strength of this impression can vary on a psychological continuum. In this he distinguishes the momentary strength of the sensation $S_{X,m}$ as a consequence of a stimulus X, and the momentary position of the various upper limits $t_{i,m}$ of the categories i (see figure 3). Both are subject to fluctuations, and will thus show a statistical spread. Thurstone assumes that these are normal distributions on the psychological continuum, thus $S_{X,m}$ comes from the distribution $N(S_X,\sigma_X)$, and $t_{i,m}$ comes from the distribution $N(t_i,\sigma_i)$.

According to the 'law of categorical judgment' the subject now places the stimulus X in the highest category i, of which the upper limit is smaller than the stimulus value, that is when $t_{i,m} \leqslant S_{X,m} < t_{i+1,m}$. Due to the fluctuations in t and S a certain stimulus will not always be judged in the same category, thus these categorical judgments also display a certain statistical distribution. Via calculation it can be shown that the categorical judgments themselves are also normally distributed when the category limits on the psychological continuum all lie equally far from their neighbouring limits.

In practice however the latter does not always apply, which indicates that in psychological terms the categorical scale is not always linear. It is now possible however to transform the distribution of the categorical judgments (i.e. the rough data on the categorical scale) in such a way that the latter is also normally distributed. The new scale is considered to correspond more closely with the psychological continuum and is called the subjective quality scale. The transformation used is based on the spread of the categorical judgments on the categorical scale: when the spread for the stimuli lying in a certain category is on average large, then the subject has not made a great discrimination in his judgments around that category. In that case the respective category is in psychological terms not so important and the width of this category must be reduced on the psychological scale.

In the same way a category which includes stimuli with on average a small spread must be enlarged. This procedure also means that the greater the capacity of a subject to reproduce and discriminate (and therefore for most stimuli a relatively smaller spread), the greater the range of his subjective quality scale.

Precisely how this transformation works and what further assumptions can be applied, is explained by Torgerson (11). He also describes the possibility of calculating a subjective scale from the spread over different subjects. In that case it is however less clear how Thurstone's 'law of categorical judgment' should now be interpreted, a problem which we consider in further detail in paragraph 3.3. Unless indicated otherwise, the data in this report are processed with the so-called 'class II, condition B' model, which means that replications occur among subjects and that as far as the spread of the categorical limits is concerned, this is assumed to be zero ($\sigma_i = 0$ for all i).

On the basis of these assumptions the subjective quality scales were calculated with the aid of the SUCINT programme (described in an IPO manual (2)). In the programme a further assumption is made, namely that the spreads of the stimulus impressions $\sigma_X$ are equated with each other for all X. Furthermore, the programme uses the calculation method of Edwards (3), in which, as anticipated, stimuli with a large spread contribute to the extending of the respective intervals, and conversely stimuli with a small spread to the reduction of the respective intervals. For technical reasons the contribution of a stimulus which occurs in only one or two categories cannot be calculated. Particularly in cases where there are few replications this can produce a rather distorted psychological scale.

The resulting subjective scales are determined but for a linear transformation (one offset and one multiplication factor which is related to the average $\sigma_X$). From this it follows that when two subjective scales describe the same psychological continuum, they can be transposed into one another by means of a linear transformation.

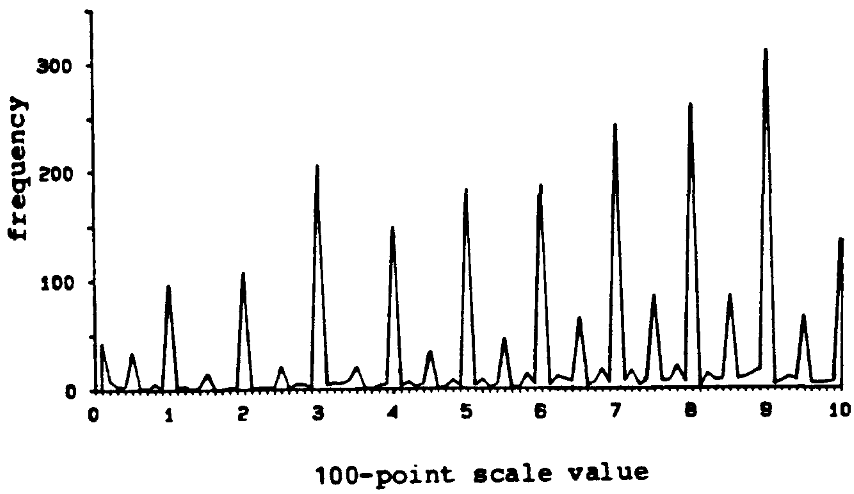



Figure 4: Histogram of categorical judgments given
The number of times that the respective report figure is mentioned is plotted for all test subjects and all stimuli together in EXP1.

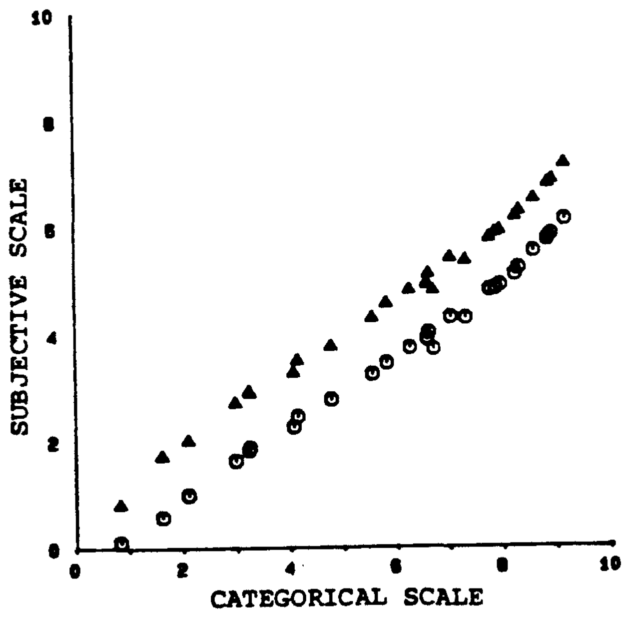



Figure 5: Subjective versus categorical scale
On the horizontal axis are plotted the average categorical scale values for the various parameter combinations. On the vertical axis are plotted the corresponding values on the subjective quality scale. There are two curves: the lower (○) relates to the subjective quality scale based on the simulated 10-point scale; the upper (▲) is based on the 100-point scale.

## 3.2 10-point versus 100-point categorical scale

The manner in which the subjects apply the 100-point scale (from 0.1 to 10.0), is considered on the basis of the judgments in EXP1.
In this experiment a number (5) of subjects made no use of the possibility of placing a figure after the decimal point. The majority (12) of the subjects mainly gave figures ending in .0 or .5. Only a few (3) made full use of the 100-point scale. A histogram of the judgments given in EXP1 (figure 4) shows these distributions.

Despite the fact that use was seldom made of the full 100-point scale, it emerged from spontaneous comments made by the subjects that the possibility of indicating a decimal was felt to be positive. This was particularly so in the case of doubts between, for example, a 7 or an 8, or between a 5.5 and a 6. The problems can then be resolved simply by choosing a category in between.

The question arises as to whether a 100-point scale in this situation has any further advantages over a 10-point scale and, what is more important, whether they both give the same results. A real test for this would be to repeat the experiment using a 10-point scale. We did not do so. In view of the fact however that the histogram of judgments in EXP1 peaks so strongly on whole numbers, we have tried to simulate this 10-point scale by means of rounding off. In this way all perceptions in the categories 6.5 up to and including 7.4 were therefore included in category '7' on the simulated 10-point scale, and mutatis mutandis for the other categories.

The respective subjective quality scales were constructed for both scales and these are shown in figure 5. It appears that the subjective scale derived from the 100-point scale barely differed from that derived from the simulated 10-point scale. The two subjective scales differ by only one constant offset, which has no further significance in view of the fact that the subjective scales are determined but for a linear transformation. The spread of the stimuli around the averaged curve is otherwise more or less identical for both scales, which likewise does not indicate an advantage in the use of a 100-point scale in preference to a 10-point scale.

Nor were any disadvantages found in the use of the 100-point scale. Due to the ease the subjects claimed to experience in using the 100-point scale, a 100-point scale was also used in EXP2.

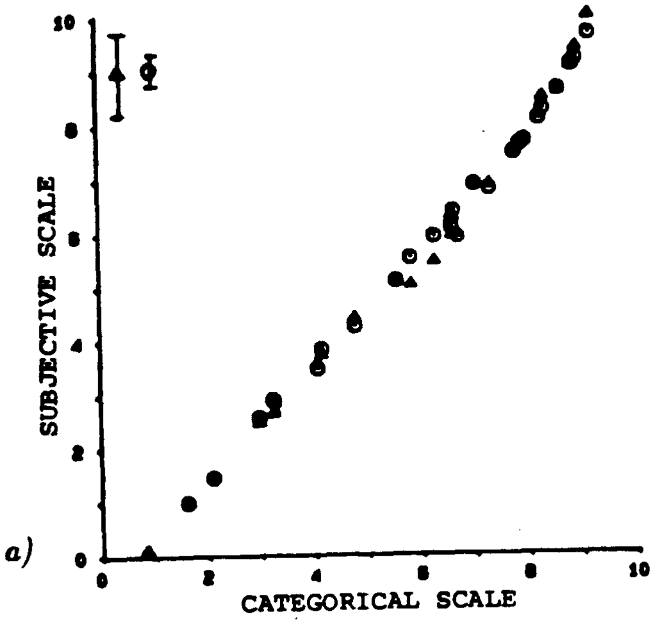
### 3.3 Spread over subjects

The subjective quality scale was determined on the basis of the spread in the categorical judgments (see paragraph 3.1). This spread can arise in several ways:

- Spread within a subject. If a subject is presented with a stimulus several times, there will be a certain spread in the categorical judgments given. It is conceivable, if this spread is considerable, that the subject is not using that portion of the categorical scale in which the respective judgments fall for making a precise discrimination. That portion is then less important in psychological terms and in the construction of the subjective scale it must be made smaller. In this case there is therefore an essential and direct relationship between the spread and the psychological scale for this subject.

- Spread over different subjects. If several subjects are presented with the same stimulus one or several times, a spread emerges because they each use the categorical scale in a different way. One subject uses, for example, only the scale values between 1.0 and 8.0, and another only those between 4.0 and 8.5. Then the size of the spread is not so much dependent on any discrimination process in giving a judgment, but more on personal differences in the use of the categorical scale. In this case it is less clear whether there is a psychological scale which is connected with the spread **over** the subjects.

The right way to construct a subjective quality scale for the 'average subject' is therefore by determining the personal subjective quality scales of a large number of different subjects and to average this out in some way or other. The method we have used until now (in paragraph 3.2) is based on the judgments of all subjects at the same time, thus on the spread **over** the test persons. Previous experiments at the IPO show that both methods produce the same result. We now wish to establish whether this also applies to the current experiments.



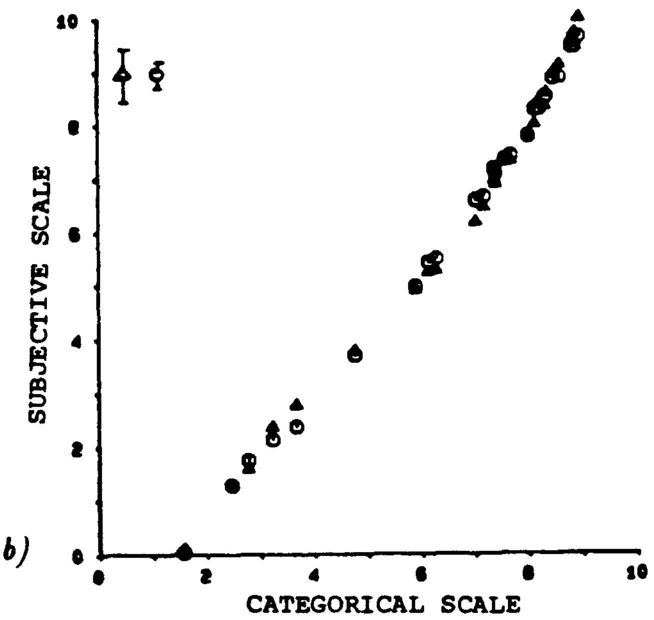




Figure 6: Comparison of T and M scales
The graph shows the transformation of the categorical scale into the subjective quality scale. On the horizontal axis are indicated the averages (over all test subjects and scenes) of the categorical judgments for the various parameter combinations. On the vertical axis the subjective quality scales are plotted: ▲ for the M scale and ○ for the T scale.

a) for EXP1

b) for EXP2.

In the construction of the personal subjective quality scales an important complication emerges: Neither in EXP1, nor in EXP2 were there different replications for all stimuli: most combinations of viewing distance, resolution, picture width and scene were presented only once to one subject. Within one subject there were therefore no real replications. Replications are however necessary for the construction of a psychological scale: for this reason we take different scenes within one subject as the replication dimension. This is the best choice because we are in the first instance not so much interested in the differences between the scenes and because, according to an analysis of variance, the different scenes appear to be one of the smallest sources of variation. The personal subjective quality scale is therefore constructed on the basis of the spread over the different scenes.

On the basis of the personal subjective quality scales the personal quality values were determined for each of the picture width/resolution/viewing distance combinations. The averages (over the subjects) of these personal quality values can be simply calculated and these are called the quality values on the M scale. The standard errors in these M values are however more difficult to determine because the personal subjective quality scales will not all have the same range (see paragraph 3.1). This means that the standard deviation of the distribution of the personal quality values for a certain parameter combination is dependent on the manner in which the ranges of the different subjects coincide. We have therefore added up a personal offset for the personal quality values of all subjects, so that all personal centres of gravity of the quality values fall on the general centre of gravity. This procedure has no consequences for the M scale itself, but it makes the standard error in the M values on average minimal.

On the basis of the judgments of all subjects for all scenes together, subjective quality values were also calculated, henceforth referred to as the T values. In order to permit a comparison between M and T values the two scales were normalised in such a way that the difference between them was minimal. The normalisation has the form of a linear transformation and is thus permitted because the subjective quality scales are in fact determined but for such a linear transformation. The results for the experiments EXP1 and EXP2 are shown in Figure 6. It appears that in both cases the M and T scales coincide well, which confirms the aforementioned IPO experiences. It is therefore acceptable to use the T scale for the study of the results in the present experiment.

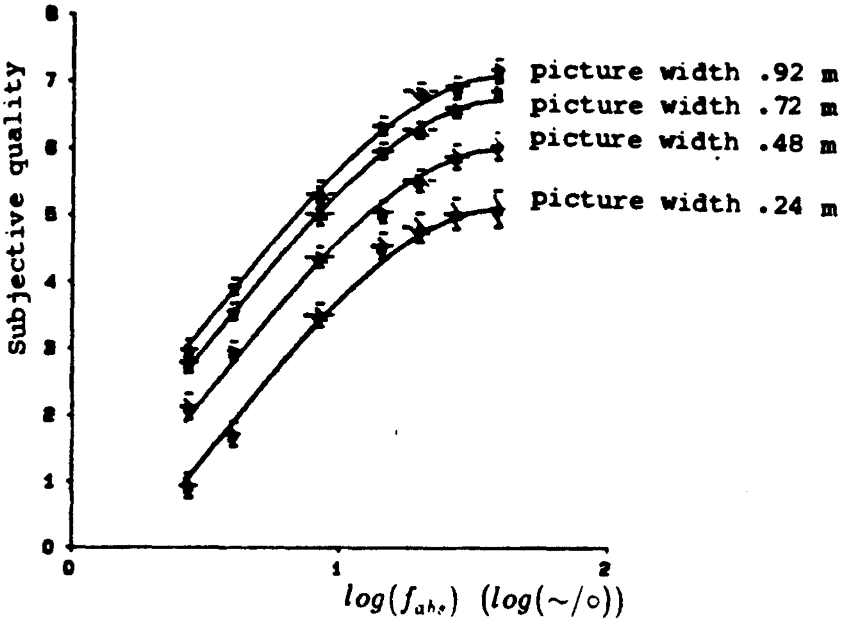



Figure 7: Subjective quality as a function of absolute resolution. The measurement points indicate the quality for a certain resolution/picture width combination. On the horizontal axis the logarithm of the resolution is plotted; the picture width is a parameter. The drawn curves have the same shape and this is calculated by a polynomial fit to the averages over the picture widths.

# 4 Results

## 4.1 Method of processing

On the basis of the results in paragraph 3.3 the following processing method was applied for both experiments: The various subjects and different scenes together are considered as the replication dimension. As variables there remain: resolution, viewing distance and picture width, these three providing a physical description of the stimulus.

The categorical judgments for these various parameter combinations were transformed to a subjective quality scale as described in paragraph 3.1. This subjective quality scale was therefore constructed on the basis of the spread over subjects and scenes. There was no deviation from this processing method unless indicated expressly in the description of the results.

## 4.2 Results of EXP1

### 4.2.1 General

In figure 7 the subjective quality values for all 28 parameter combinations are plotted as a function of the absolute resolution (expressed in ~/o, see paragraph 2.3), and with the picture width as the parameter. It appears that the subjective quality for all four picture widths changes in the same way with the resolution: the four curves only differ by a constant offset. The **shape** of the curve is determined by averaging the four measurement points for a single resolution, and fitting a polynomial to these average measurement points. It appears that a third-order polynomial fit is well-suited for this. The dependence of the subjective quality on the resolution can therefore be interpreted as follows: For low resolutions the quality rises rapidly with increasing resolution, but there is a saturation at approximately 25 ~/o, which lies in the order of magnitude of the limitations of human vision. In the case of resolutions higher than 25 ~/o improvements in the resolution can therefore be barely perceived any longer by the eye and the curve becomes flatter and finally horizontal.

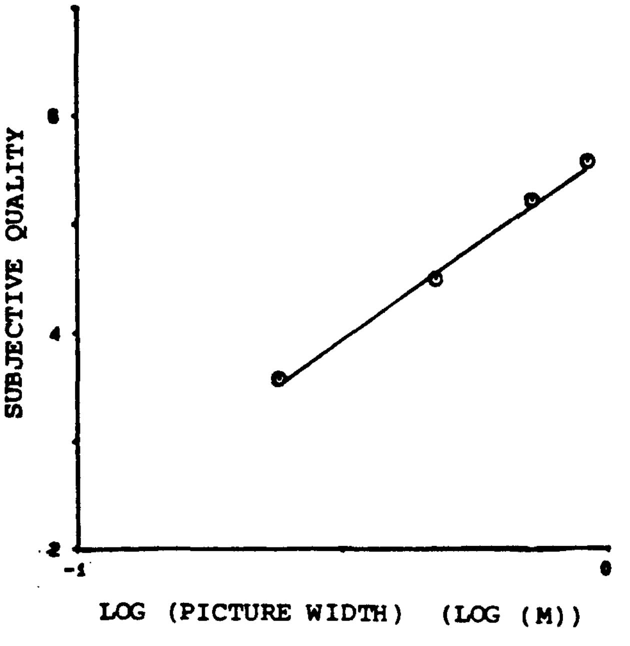



Figure 8: Quality averages as a function of the picture width.
The values plotted are the averages over all resolutions at a fixed picture width. They are indicated as a function of the logarithm of that picture width.

The vertical displacement of the curves as a result of the variation in picture width can be studied on the basis of the quality values averaged over all resolutions. Figure 8 shows that there is a linear relationship between the logarithm of the picture width and these averages.

The two descriptions can be summarised in the following formula:

$$Q(b,f_{abs}) = 3.460\log(b) + 1.199 + 3.827\log(f_{abs}) + 2.217(\log(f_{abs}))^2 - 1.416(\log(f_{abs}))^3,$$

when $f_{abs}$ 40 ~/o, and for higher resolutions the saturation value is simply calculated. In the formula the resolution $f_{abs}$ is expressed in ~/o, and the picture width **b** in metres ($0.24 \leqslant b \leqslant 0.92$). The first two terms describe the influence of the picture width on the subjective quality; the following terms give the third-order polynomial adjustment for the dependence of quality on the absolute resolution.

It can be gathered from both the aforementioned formula and the graph that picture width and resolution influence the subjective quality independently. In other words: resolution and picture width show no interaction. This latter claim is not however substantiated by a variance analysis on the rough data: in this (with subjects and scenes as the replication dimension) a significant interaction is found ($F(18/1782) = 3.63$, $P = 0.001$). Even if all replications are considered as completely independent (i.e. the dimensions of resolution and picture width are 'nested' within the replication dimension), the interaction remains significant at 5% level: $F(18/2772) = 1.61$, $P=0.05$. An explanation for this may lie in the fact that the variance analysis is applied to the rough, not psychologically scaled data, while the graphs in figures 7 and 8 are based on the subjective quality scales.

It would be better to apply the analysis of variance to the total of all judgments, which in that case must be corrected in such a way that they describe the subjective quality scale. One way to approach this is by making a polynomial fit of the curve which transforms the category scale

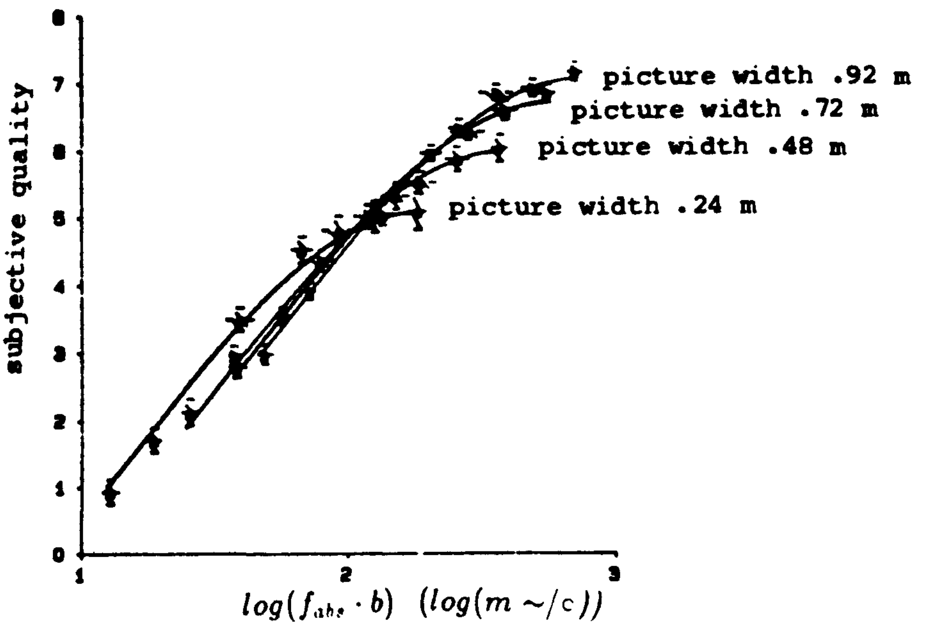



Figure 9: Subjective quality as a function of relative resolution
Measurement points and curves are the same as in figure 7, but on the horizontal axis the parameter $f_{abs}$.b is now plotted. This parameter is for a fixed viewing distance proportional to the relative resolution.

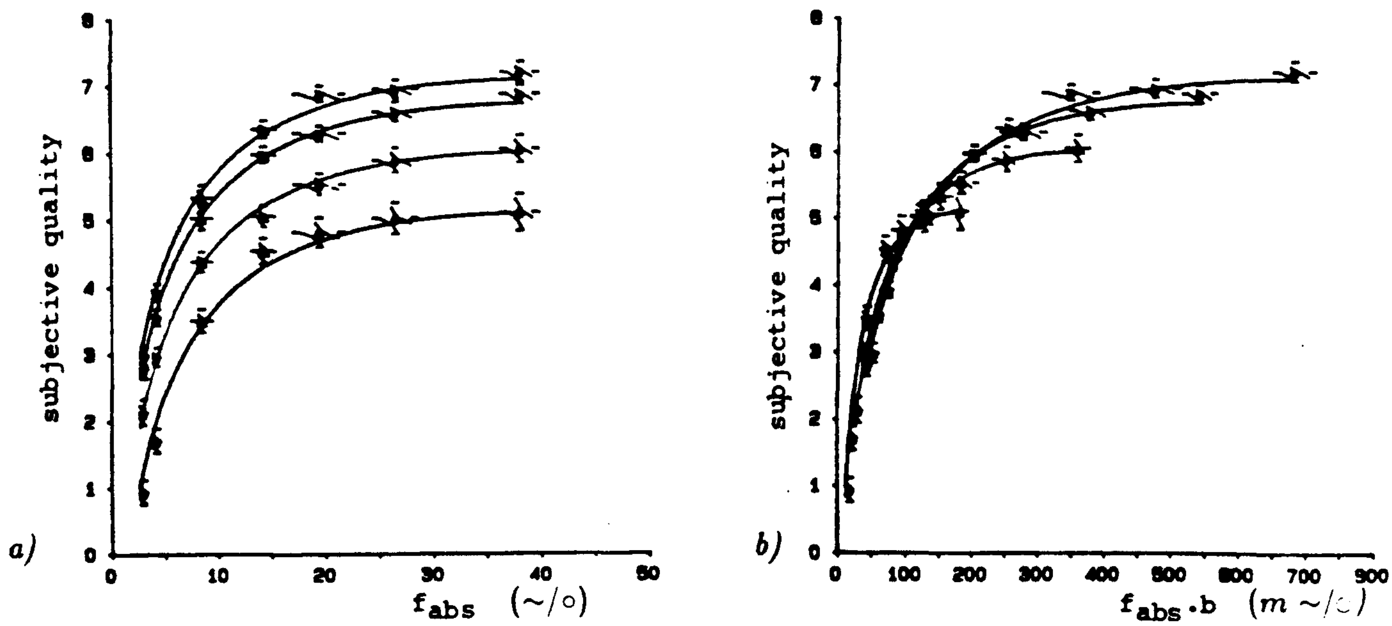



Figure 10: Subjective quality as a function of resolution
The two graphs are the same as those in figures 7 and 9, but the resolutions on the horizontal axis are now plotted linearly.
a) Subjective quality as a function of absolute resolution.
b) Subjective quality as a function of relative resolution.

into the subjective scale (this is in fact the T scale in figure 6), and by subsequently converting the rough categorical judgments with the aid of this polynomial fit into judgments on the subjective quality scale. It in fact appears that if a analysis of variance is applied to these slightly corrected data, the interaction between resolution and picture width is no longer significant ($F(18/2772)=1.51$, $P=0.08$).

**4.2.2 Deduced observations**

In the previous paragraph a formula was found which describes the measurement results well. With the aid of this formula it is now possible to consider the results from various points of view.

First of all a single parameter is required which gives a good description of the quality. The relative resolution is a possible candidate. The relative resolution is related to the width of the picture and is defined as the maximum number of periods fitting into the picture width; the corresponding unit is consequently periods. For a fixed viewing distance a the relative resolution is proportional to the absolute resolution multiplied by the picture width: $f_{rel}=360/2\pi . f_{abs} . b/a$.
The concept of relative resolution is closely related to the concepts of 'number of pixels' and 'bandwidth' in the video world. Figure 9 shows that the relative resolution is indeed a better predictor of subjective quality than absolute resolution or the picture width alone. This is understandable because this relative resolution is a measure of the quantity of information which is transmitted.
The description in terms of relative resolution is not however completely satisfactory because the four curves do not completely coincide.
Calculation shows that this would happen for the unsaturated portion if the parameter $f_{abs}{}^{1.4} . b$ were to be plotted on the horizontal axis. It is however difficult to give this product an interpretation.

The graphs in figures 7 and 9 are placed in a different light if we take a linear axis instead of a logarithimic resolution axis, as was done in figure 10. The saturation is now shown more clearly and consequently the effects resulting from the bandwidth also emerge more distinctively.

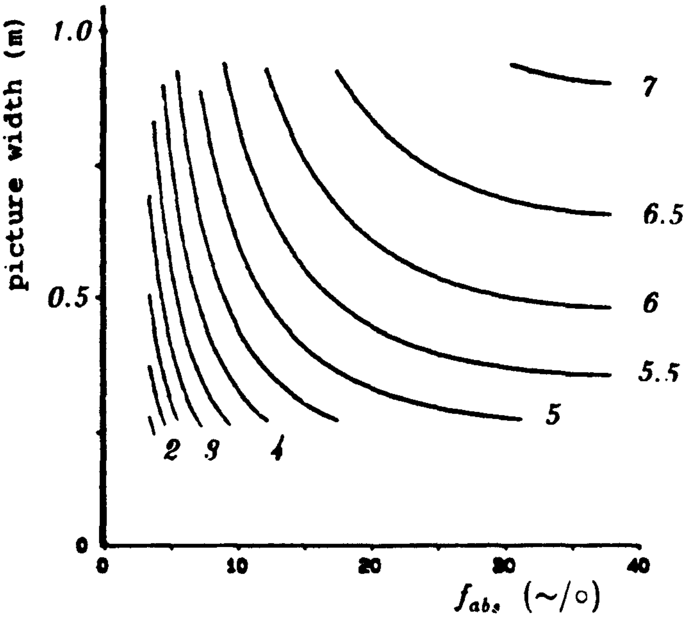



Figure 11: Isoquality curves
The lines in the graph connect the combinations of picture width and resolution which produce the same quality. The figures alongside the curves indicate how high that quality is.

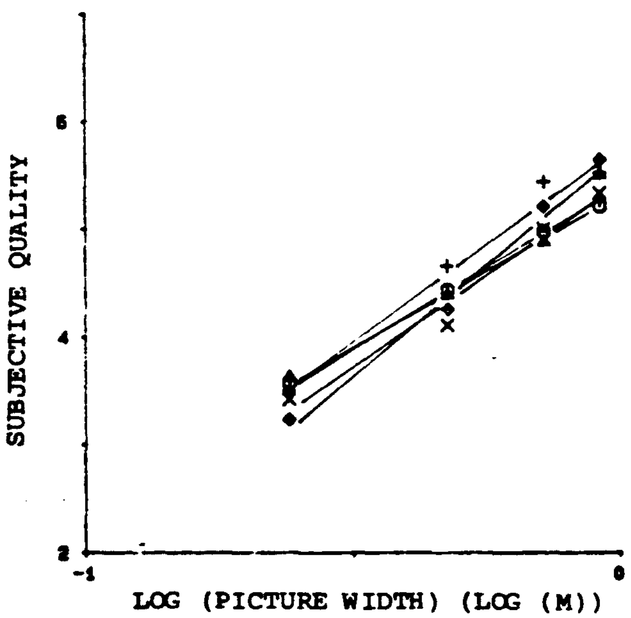



Figure 12: Scene dependence of the subjective quality
The averages over the resolution are indicated as a function of the logarithm of the picture width, in the same way as the graph in figure 8. In this case, however, the different scenes are kept separate:

○) 'ropes' scene
△) 'Mint tower' scene
+) 'Graffiti wall' scene
×) 'terrace' scene
◇) 'Thielke' scene

A third alternative manner of interpreting the results is on the basis of isoquality curves, as shown in figure 11. These are calculated by setting the quality in the formula at a fixed value and observing how resolution and picture width must exchange their positive influences for this quality value to remain constant. From this graph it is possible to establish the best way of working on a quality improvement: this being in the direction perpendicular to the isoquality curves. Whether that comes down to an increase of the size or of the resolution is dependent on the initial situation.

#### 4.2.3 Different scenes

Up to now the scenes dimension has always been used as the replication dimension, based on the assumption that the scene content has no influence on the evaluations of the subjects. In this paragraph it is considered to what extent this assumption is justified.

From a variance analysis on slightly corrected categorical judgments (they were converted with the aid of polynomial fit from a categorical to a quality scale, see also paragraph 4.2.1) it appears that there is no interaction between the resolution and the scenes ($F(24/2660)=0.83$, $P=0.7$). The interaction between scenes and picture width is however important ($F(12/2660)=3.9$, $P< 0.001$), and the scene main effect too is significant ($F(4/2660)=9.21$, $P< 0.001$).

In order to establish in what way these effects manifest themselves, the rough data are once again scaled with the aid of the SUCINT programme, but in this case with only the subjects as the replication dimension, and therefore with 140 different stimuli. Then the averages are taken of the resolution per scene and per picture width. This averaging is permitted because it has been shown from the variance analysis that there is no interaction between scenes and resolutions. These average quality values however show clearly the scene interaction and main effects.

From figure 12 it can be seen that for each scene the relationship between these averages and the logarithm of the picture width can be described by a linear relationship. This is also confirmed by the respective correlation coefficients which are all five higher than

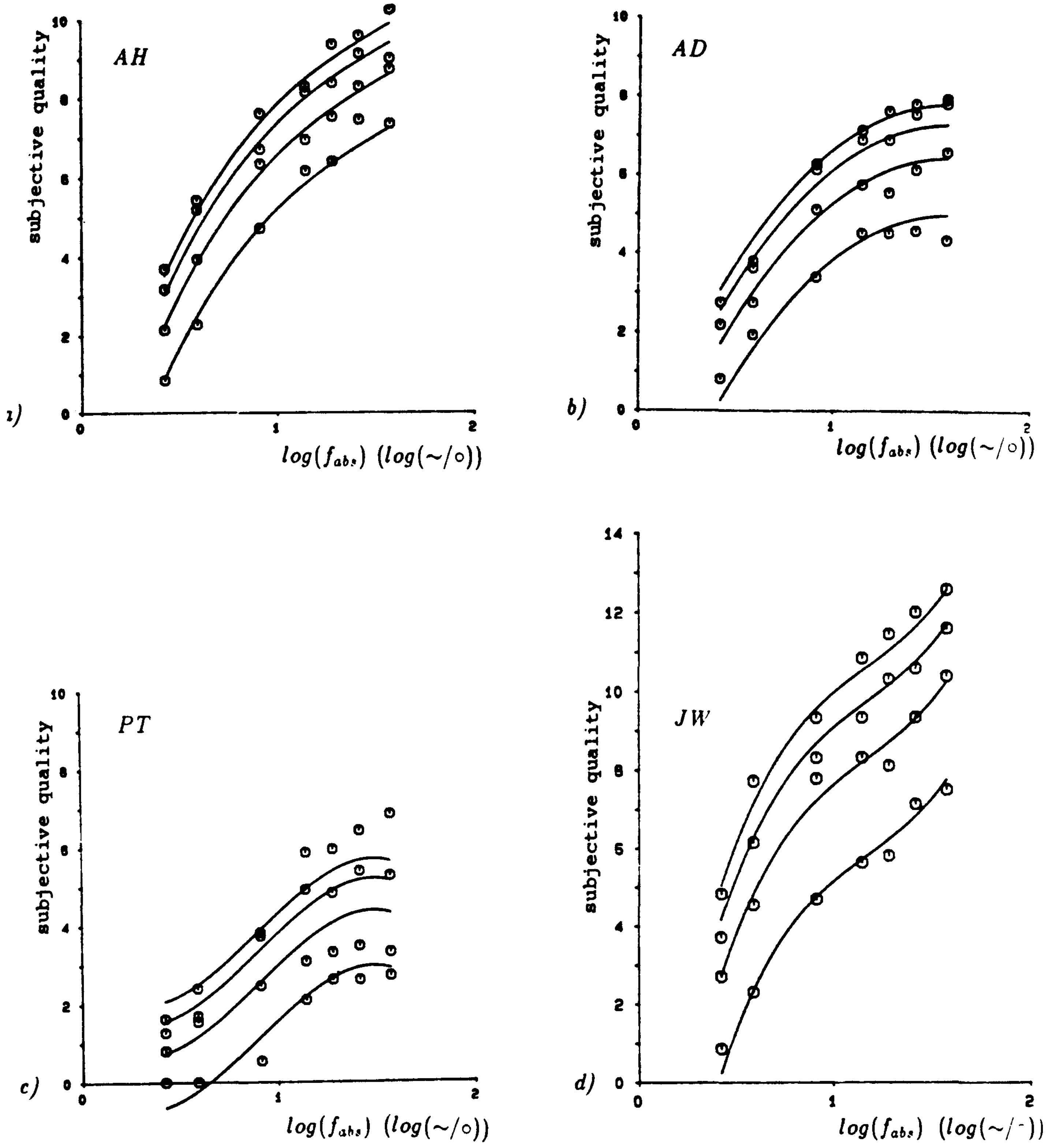



Figure 13: Personal subjective quality
These graphs indicate the subjective quality values as a function of the resolution, with as the parameter the picture width, in the same way as figure 7. The shape of the curves drawn is also calculated as a polynomial fit to the averages over the picture width. Each graph is based on the judgments of only one test subject:

a) test subject AH, vision 2.00
b) test subject AD, vision 2.50
c) test subject PT, vision 2.00
d) test subject JW, vision 2.00.

0.984. A difference in slopes does indeed emerge: these are smaller than average for the 'ropes' and 'Mint tower' scenes, and on the other hand relatively high for the 'Thielke' scene. Generally speaking, these differences are not however dramatic, and it can be stated that the linear relationship found in paragraph 4.2.1 between quality and the logarithm of the picture width describes the general trend well.

### 4.2.4 Different subjects

In paragraph 3.3 it has already been shown that the subjective quality scale for the total of all rough data is the same as the scale which emerges from the averages of all personal subjective quality scales. In this paragraph these personal subjective quality scales are considered in greater detail. The purpose of this is to gain insight into the possible differences between subjects.

For each subject a personal psychological scale was calculated on the basis of the spread over the different scenes. The subjective quality values found are reproduced in figure 13 as a function of the resolution for four subjects. For each subject a polynomial adjustment was made of the average of the picture width as a function of the resolution, in the same way as was done in paragraph 4.2.1 for the common quality data.

It can be deduced from the graphs that for three of the four subjects (AH, AD and JW) the form of this curve is not in fact dependent on the picture width. For subject PT the deviations are however greater. The reason for this might be that the spread in the quality judgments for the different scenes in the case of PT is so small that it is difficult to construct a personal subjective quality scale for this subject (see paragraph 3.1). This might also serve to explain why the subjective quality scale of PT is much smaller in size than that of the other subjects.

The overall form of the drawn curves corresponds qualitatively with that for all the subjects together (figure 7): a rising portion for low resolutions and a gradual saturation for higher resolutions. Subject JW (vision 2.0) does not however fit in with this picture, because the subjective quality continues to increase even with the highest resolutions. The high sensetivity can be explained by the fact that subject JW is the one who set up the experiment and as a result became trained in distinguishing different resolutions.

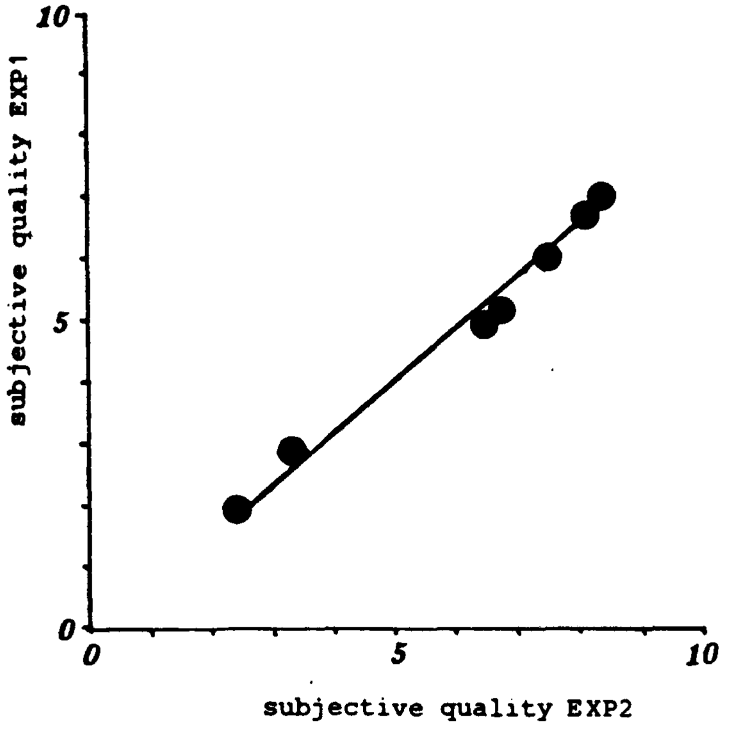



Figure 14: Comparison of the psychological scales of EXP1 and EXP2. On the horizontal axis are plotted the subjective quality values in EXP1 of a number of parameter combinations presented in both experiments. On the vertical axis are indicated the subjective quality values in EXP2 for these same parameter combinations. The drawn line gives the best linear fit to these values (correlation 0.993).

Nevertheless, the graph concerned does pose the question to what extent the saturation of the curve in the case of other subjects is a consequence of the restrictions of the visual system, in view of the fact that subject JW clearly shows that it is indeed possible to make a distinction between these resolutions.

An answer to this question can perhaps be given by classifying the subjects according to their vision ( $\in$ 1.5, 2.0 or $\geqslant$ 2.25). For all three groups the subjective scales were determined, and with the aid of these the shape of the curves by means of polynomial fits to the averages of the picture width. For these curves a critical resolution was defined as that resolution at which the oblique asymptote (for low resolutions) cuts through the horizontal distinction between these resolutions (for saturated quality). One would expect this critical resolution to increase with the vision of the subjects. This is however not found, while a non-significant trend in the critical resolutions points in precisely the opposite direction. It appears that the critical resolution, due to a relatively large inaccuracy in combination with the small differences in vision, is not able to show a dependence of the eyesight.

**4.3 Results of EXP2**

**4.3.1 Comparison with EXP1**

In paragraph 3.1 it has already been explained that when two psychological scales describe the same psychological continuum, they can be transposed into one another by a linear transformation. To what extent the subjective quality scales of EXP1 and EXP2 satisfy this can be established on the basis of a set of eight parameter combinations, which was presented in both experiments. In figure 14 the quality values of these parameter combinations in EXP2 are plotted against those in EXP1. The graph shows that there is in very close approximation a linear relationship between the two data sets (correlation 0.993). It can be concluded from this that the two subjective scales are in conformity with each other and describe the same psychological continuum. Moreover, the linear fit found offers the possibility of transforming the quality values from EXP1 in such a way that they can be directly compared with the results in EXP2.

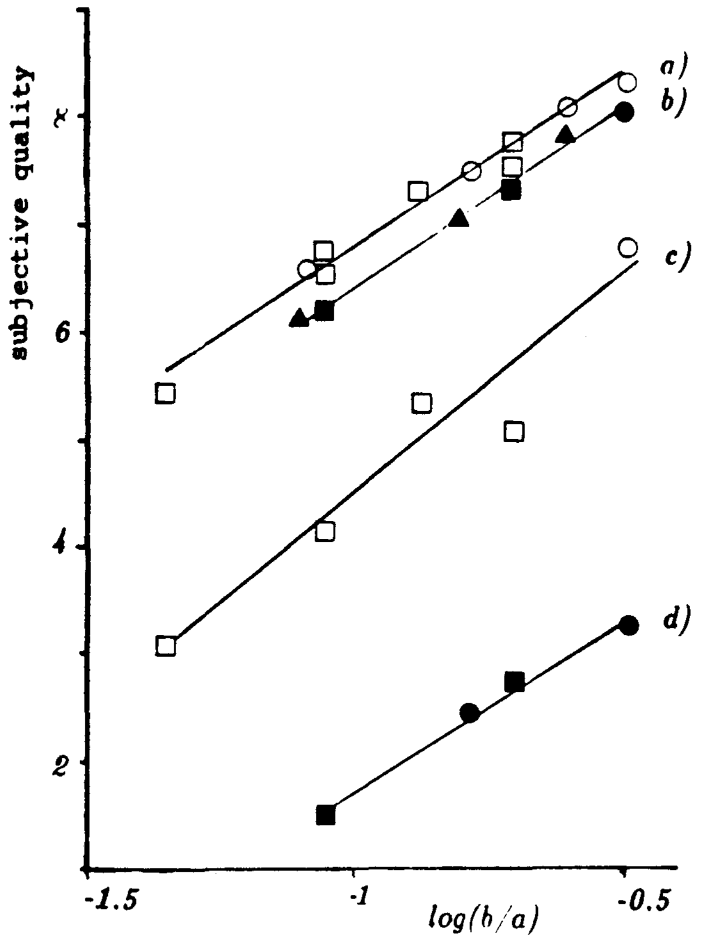



Figure 15: Subjective picture quality as a function of the picture angle. The subjective quality values on the vertical axis are plotted as a function of log(b/a) on the horizontal axis. The quotient of picture width **b** and viewing distance **a** is proportional to the picture angle ϕ. There are four different data sets, corresponding to four different resolutions:

a) resolutions greater than 33 ~/o,

b) resolutions between 23 and 28 ~/o,

c) resolutions between 8.6 and 8.7 ~/o,

d) resolutions between 2.6 and 2.7 ~/o.

Different symbols are also used to permit a distinction to be made between results for different viewing distances:

O,●) viewing distance 2.9 m,

△,▲) viewing distance 3.9 m,

□,■) viewing distance 5.4 m.

**4.3.2 General**

Because the experimental configuration of EXP2 is not completely crossed (see paragraph 2.2), it is not possible to apply a variance analysis to the total of all data. In order to nevertheless gain an idea of the interactions to be anticipated, a selection has accordingly been made: only stimuli with a picture width of 0.92 m or 0.48 m, a viewing distance of 2.9 m or 5.4 m, and resolution which is maximum or approximately 2.6 ~/o, were considered. The rough categorical judgments for these eight parameter combinations were transformed to a subjective scale with the aid of a polynomial fit of the relationship between the categorical scale and T scale of EXP2 from figure 6 (the procedure followed for this is the same as that for EXP1, described in paragraph 4.2.1.). A variance analysis of these eight slightly corrected parameter combinations produced significant main terms, but no significant interactions: for all four possible interactions the following applies: $F(1/632) \leqslant 0.99$, $P \geqslant 0.32$.

On the basis of this prediction that there are no interactions to be anticipated, and of the results known from EXP1 relating to picture width and resolution, it is now particularly interesting to observe the way in which the main term effect manifests itself as a consequence of the viewing distance. Figure 15 shows that for all resolutions the influences of the viewing distance and picture width can be summarised in an effect of picture angle. This picture angle $\phi$ is calculated as $360/2\pi$. **arctan** (b/a), and is proportional to the size of the picture on the retina. For each resolution the subjective quality appears to depend linearly on the logarithm of this picture angle: each of the four straight lines in the graph has a correlation higher than 0.984. Furthermore, the adjustments run more or less in parallel (slopes between 3.3 and 4.3, with an average of 3.593), which reflects the absence of any interaction between resolution and viewing distance or picture width. Nor do there appear anywhere in the graph any clear systematic effects in consequence of the viewing distance, which corresponds with the prediction that there is no interaction to be anticipated between viewing distance and picture width.

The main effects in consequence of viewing distance and picture width are therefore adequately described with the aid of the picture angle; what is still needed is an adequate description of the influence of the resolution on the subjective quality. One way of giving expression to these influences is by correcting the quality data for the picture angle effect already known:

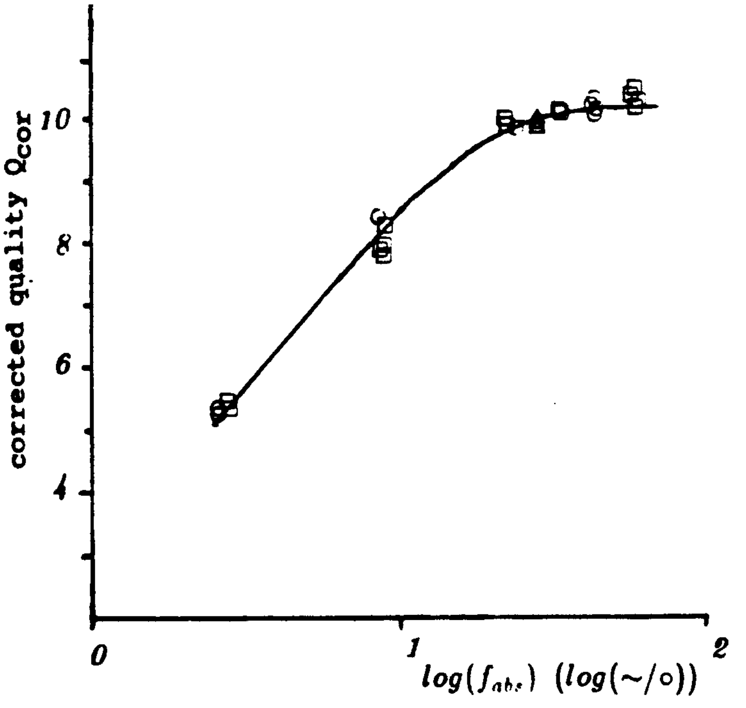



Figure 16: Corrected quality as a function of resolution
The quality values plotted on the vertical axis are calculated from the original quality values by correcting for the influence of the picture angle already known: $Q_{cor} = Q-3.593 \times log(b/a)$. In order to make a distinction between the viewing distances different symbols were used: ) viewing distance 2.9 m, ) viewing distance 3.9 m, ) viewing distance 5.4 m. The drawn curve is not a fit to the data, but based on the curve shape as found in EXP1 (see text).

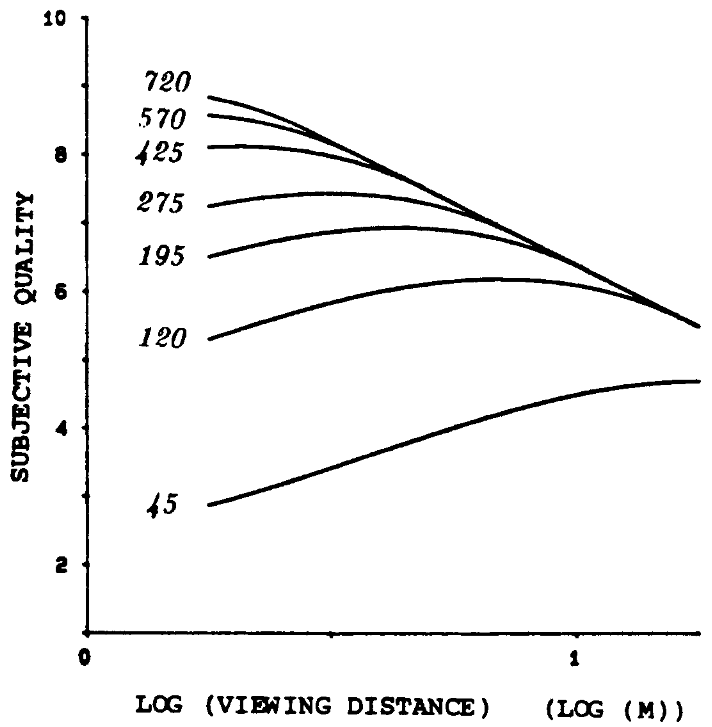



Figure 17: Subjective quality as a function of viewing distance
The drawn curves are extrapolations on the basis of the quality formula of paragraph 4.3.2. They are calculated in the case of different relative resolutions varying between 720 and 45 periods, as is indicated next to the curves, and for a picture width of 0.92 m. For smaller picture widths the whole graph shifts downwards.

$$Q_{cor} = Q - 3.593.\log(b/a).$$

These corrected quality values are plotted as a function of the resolution in figure 16. They all appear to lie with a high level of precision on a single curve, of which the shape corresponds with the results of EXP1. The curve drawn in the graph is calculated on the basis of the polynomial portion of the quality formula for EXP1 in paragraph 4.2.1. This is then converted to the subjective scale of EXP2 on the basis of the linear relationship between EXP2 and EXP1, as described in paragraph 4.3.1. Figure 16 shows not only once again that the results of EXP1 and EXP2 are in agreement with each other, but also that for the different viewing distances the subjective quality depends in the same way on the resolution.

The results of EXP2 can be summarised in the following formula:

$$Q = 3.593\log(b/a) + 2.886 +$$
$$4.608\log(f_{abs}) + 2.669(\log(f_{abs}))^2 - 1.705(\log(f_{abs}))$$

Subjective quality appears to be determined by two parameters: effects in consequence of the picture angle are represented by the first term and the influence of the resolution is expressed by the last three terms.

#### 4.3.3. Deduced observations

With the aid of the formula at the end of the previous paragraph it is possible to consider the results of EXP2 in an alternative manner, as was also done in paragraph 4.2.2 for EXP1. It is therefore also possible for EXP2 to reproduce the data as a function of the relative resolution $f_{rel}=360/2\pi.f_{abs}.b/a$.

The results of EXP2 essentially provide new insight as regards the dependence on viewing distance of the subjective quality. For a picture with a given width and relative resolution the viewing distance appears to influence the quality in two ways: firstly, an increasing viewing distance has a negative influence on the quality, because the picture angle in that case becomes smaller. Secondly, it also plays a role in determining the absolute resolution (see paragraph 2.3) and in such a way that with increasing viewing distance the quality improves. The two effects together ensure the existence of an optimal viewing distance, which can also be taken from figure 17. From the quality formula it can be deduced that the optimal viewing distance is always chosen in such a way that the absolute resolution equals 16 ~/o. This has the result that the optimal viewing distance is determined by the 6 dB cut-off frequency on the screen (expressed in ~/m), and is independent of the picture width. The picture width does however play a part, together with the 6 dB cut-off frequency on the screen, in determining how great the subjective quality will be in the case of this optimal viewing distance.

# 5 Conclusions and discussion

## 5.1 Comparison of the two experiments

The results of the experiments EXP1 and EXP2 appear to confirm one another in most cases with a very high level of accuracy. One of the main points of correspondence is that both experiments describe the same psychological continuum (paragraph 4.3.1), despite the fact that for the most part different subjects took part in the experiments. This appears to fully confirm the presupposition that there is such a thing as a universal average psychological continuum.

Not altogether separate from this is the excellent qualitative correspondence between the quality values found. As regards the shape of the curve (quality as a function of resolution) we find even in respect of quantity the same results in both experiments (paragraph 4.3.2). The shape of the curve consists of a rising quality with increasing resolutions, which partially saturates at resolutions above 25 ~/o. Presumably this saturation is to do with the limitations of the human eyesight (60 ~/o), but this cannot be shown with the aid of the existing data.

The way in which the quality depends on the picture width **b** is also the same in both experiments: there is no interaction with the resolution and there is a linear relationship between the quality and the logarithm of the picture width. The slope of this relationship is however in both experiments **not** in the same proportion to the shape of the quality curve as a function of the resolution. This is manifested in the different parameters which are required for the unsaturated portions of the quality curves to coincide with each other: $f_{abs}^{1.4}.b$ in EXP1 and $f_{abs}^{1.6}.b$ in EXP2 (paragraphs 4.2.2 and 4.3.3.). An explanation for this has not been found, nor can the use of the 'Mint tower' scene in EXP1 (and not in EXP2) be considered as a possible reason. The fact that this scene has a smaller slope than average in the linear relationship between quality and picture width (paragraph 4.2.3) should give rise to precisely the opposite effect.

Despite this small discrepancy the general conclusion remains that picture angle $\phi = 360/2\pi.b/a$ and resolution $f_{abs} = 2\pi/360.f_{rel}.\ a/b$

both influence the subjective quality without interaction. This is expressed in the following formula:

$$Q(f_{abs},ba) = 1/K \cdot \log(b/a) + 0.80\log(f_{abs}) + 0.46(\log(f_{abs}))^2 - 0.29(\log(f_{abs}))^3$$

This formula has as far as possible been stripped of non-essential constants, which is possible because the subjective scale is determined but for a linear transformation (paragraph 3.1.). The constant K in this case has a value of around 1.5 (between 1.4 and 1.6).

## 5.2 Size-constancy effect

Van der Zee and Boesten (14) report a size-constancy effect in the judgment of the subjective quality. In the experiments in question the picture width b, viewing distance a and luminance were varied for slides with a very high resolution. It appeared that in the case of a fixed luminance the subjective quality is dependent on $\log(b^2/a)$, and can be described by this parameter **alone**. Apparently the size of the picture on the retina, proportional to b/a, and the size of the picture in reality play an equally important role here. The fact that the picture width itself is also of direct influence is related by Van der Zee and Boesten to the known size-constancy effect (see for example Gregory (4)). Hatada (5) also finds such a relationship in an experiment in which only picture width and viewing distance were varied.

In the results of EXP2 however there is no question at all of size-constancy effect: only the picture angle $\phi$ proportional to the size of the image on the retina influences the picture quality of pictures with a very high resolution. The reason for this lack of conformity should probably not be sought in differences in the experimental configurations. As far as can be established, EXP2 and the measurements of Van der Zee and Boesten were carried out under identical conditions. The sole difference, namely that in EXP2 there was extra information available on the distance in the form of identical objects, could only serve to contribute to the emergence of a size-constancy effect.

It is more probable that the differences arising are due to the perceptive dimensions in which the variation of the stimuli takes place. In EXP2 the resolution is varied over a broad range, as a result of which it is essentially the percept 'sharpness' which is addressed. The sharpness impression will therefore play an important part in determining the quality judgment. In the experiments of Hatada and Van der Zee and Boesten the resolution is however kept at maximum, so that the subjects will also be less inclined to use the sharpness impression as an element of the quality criterion.

This explanation leads however to the conclusion that the quality criterion used by the subjects is influenced by the choice of stimulus. This would however considerably restrict the possibility of mutual comparison and the application of the results of this type of experiment. Research into the stability of the quality criterion therefore appears to be highly desirable.

### 5.3 Optimal viewing distance

Jesty (7) describes an experiment in which the subject is confronted with a projected scene and he is asked to place his chair at that position at which he prefers watching the scene. This is repeated for several picture widths and for various degrees of defocusing. It appears that the quotient of the optimal viewing distance $a_{opt}$ found and the picture width is constant for different picture widths and dependent on the relative resolution: $b/a_{opt} = C(f_{rel})$. Although Jesty nowhere gives an explicit definition, it appears from the article that he considers this quotient as a measure of the quality, which he otherwise consistently refers to as 'sharpness'.

In paragraph 4.3.3 we calculate on the basis of the results of EXP2 that the optimal viewing distance is found for a fixed resolution of 16 ~/o, independent of the picture width. It can be derived from this that for the optimal viewing distance at different picture widths and relative resolutions, it applies that $a_{opt} = 16.360/2\pi.b/f_{rel}$. This fully corresponds with the findings of Jesty.

It is now also possible to calculate the subjective quality for this optimal viewing distance: it appears that this maximum achievable quality is not dependent on the picture width and depends linearly on the logarithm of $f_{rel}$. This indicates that Jesty is wrong in suggesting that the (maximum achievable) quality of a display is proportional to $b/a_{opt}$, although he rightly concludes that it is independent of the picture width.

In a second experiment of this type Jesty varies the picture width for a number of resolutions-on-the-screen ($f_{6dB}$) by reducing the size of the projection screen. For small picture widths parts of the picture therefore disappear. Jesty considers that for a fixed value of $f_{6dB}$ the optimal viewing distance varies slightly with the picture width: $b^{0.19}/a_{opt} = C(f_{6dB})$. The influence of the picture width therefore appears to be very small, although the results of EXP2 predict a total independence.

Nor do the results of Jesty's second experiment correspond with those of his first experiment. The discrepancy can perhaps be explained by the different ways in which the picture width is varied. In EXP2 and in Jesty's first experiment the total quantity of information originally present is kept constant (therefore the entire scene is projected), whereas in his second experiment he changes this quantity together with the picture width. It therefore appears useful to consider whether the reduction of the total quantity of information available does indeed lead to the optimal viewing distance no longer being solely dependent on the relative resolution, and if so, what psychological consideration is at the root of this.

### 5.4 Relationship with TV and High Definition TV

One could ask to what extent the results described can be directly translated into conclusions for television. There are in fact two important differences between the stimuli presented and the TV in the living room. The first is that in the experiment only still slides were used, whereas television pictures move. To date it is unknown in the literature how the percept of sharpness behaves in the case of moving pictures. On the basis of the fact that the emotional involvement is less in the case of still pictures and that there is more time to take in the scene in detail, one would expect the sensitivity to changes in resolution to be less in the case of moving pictures. In that case the experiments described should be considered as a type of 'worst-case' approximation, and directly applicable to the - not always negligible - portion of static television pictures.

A second difference lies in the way in which a certain resolution arises. In the experiments described the resolution is equally great in the horizontal and vertical direction, and in both directions the picture has an

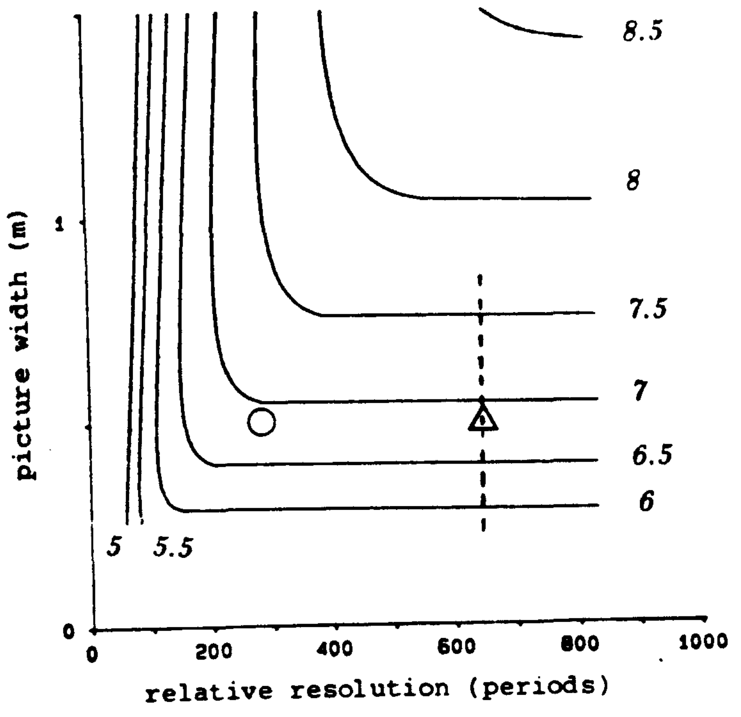



Figure 18: Isoquality curves
The drawn curves connect combinations of picture width and relative resolution which produce the same quality. They are calculated for a viewing distance of 4 m. A number of points is drawn on the graph: O gives approximately the present-day TV, and Δ and the dotted line through it indicate the possibilities of HDTV.

identical information structure. This also applies to the horizontal direction of television pictures, but in the vertical direction the picture information is made discreet in the distinct scan lines. Although it is also possible for TV pictures to have the same resolution in both directions, the nature of the information reproduced is different (horizontally analog, vertically digital). It is probable that the manner of information structuring also influences the sharpness percept. Continued research is required to show to what extent this can be of importance.

The direct translation of the results of this research in terms of television signals is thus subject to a number of reservations. Despite this some effects are so great and so distinct that a predictive value can be attached to them. One example of this is the much-discussed conversion from TV to HDTV, which is defined as a doubling of the relative resolution. Figure 18 shows the overall positions of TV and HDTV. It can be observed from the graph that this improvement will have absolutely no effect in many living rooms if it is not coupled with an increase of the picture width.

**References**

(1) M. Born, E. Wolf, Principles of Optics, Pergamon Press, London, 19

(2) R. Boesten, Categorie schaling: the method of successive interval scaling, SUCINT programme, IPO Manual no.55.

(3) A.L. Edwards, Techniques of Attitude Scale Construction, Appleton-Century-Crofts, New York, 1957.

(4) R.L. Gregory, Eye and Brain; the Psychology of Seeing, Weidenfeld and Nicolson, London, 1966.

(5) T. Hatada, H. Sakata, H. Kusaka, Psychophysical analysis of the 'sensation of reality' induced by a visual wide-field display, SMPTE J. 89(1980)560-569.

(6) G.C. Higgins, Image quality criteria, Journal of Applied Photographic Engineering 3(1977)53-60.

(7) L.C. Jesty, The relation between picture size, viewing distance anpicture quality, Proc. IEE 105B(1958)425-439.

(8) R. Röhler, Informationstheorie in der Optik, Wissenschaftliche Verlagsgesellschaft MBH, Stuttgart, 1967.

(9) H.L. Snyder, Image quality and observer performance, in: Perception of Displayed Information, L. Biberman, editor, Plenum Press, New York 1973.

(10) H.L. Task, A comparison of several display image quality measures, Proc. SID 19(1978)113-119.

(11) W.S. Torgerson, Theory and Methods of Scaling, John Wiley and Sons New York, 1958.

(12) J.H.D.M. Westerink, Perceptief gewogen spatiale resolutie, IPO-rapport no.587, 1987.

(13) E. van der Zee, M.H.W.A. Boesten. The influence of luminance and size on the image quality of complex scenes, IPO Annual Progress Report 15(1980)69-75.

(14) E. van der Zee, M.H.W.A. Boesten, A.L. Duwaer, The influence of 'size constancy' on the subjective evaluation of image quality, IPO Annual Progress Report 18(1983)61-66.

Appendices

**A Instructions for the subject**

For EXP1:

Geachte proefpersoon,

U werkt mee aan een experiment betreffende beeldkwaliteit. Doel van het experiment is, vast te stellen, hoe de door mensen ervaren kwaliteit afhangt van een aantal factoren, zoals scherpte en grootte van het beeld.

In dit experiment worden een aantal dia's vertoond. Zij verschillen in scherpte, grootte en voorstelling.

Het is de bedoeling, dat U elke dia op zijn kwaliteit beoordeelt, door een rapportcijfer voor die kwaliteit te geven. Hierbij moet U uiteraard zo min mogelijk op de voorstelling zelf letten.
Dus:
10.0 = uitstekend (maximaal)
0.1 = erbarmelijk (minimaal)

N.B.:
U mag dus een cijfer achter de komma geven.

In totaal krijgt U 147 dia's te zien, gedurende 15 sekonden elk.
De eerste zeven dia's zijn proefdia's, en zij geven een een goed beeld van de kwaliteits-variatie van de resterende 140 dia's.

Als U nog vragen hebt, stel die dan gerust aan de proefleider. Anders, veel succes!

For EXP2:

Geachte proefpersoon,

U werkt mee aan een experiment betreffende beeldkwaliteit. Doel van het experiment is, vast te stellen, hoe de door mensen ervaren kwaliteit afhangt van een aantal factoren, zoals scherpte en grootte van het beeld.

In dit experiment worden een aantal dia's vertoond. Zij verschillen in scherpte, grootte en voorstelling.

Het is de bedoeling, dat U elke dia op zijn kwaliteit beoordeelt, door een rapportcijfer voor die kwaliteit te geven. Hierbij moet U uiteraard zo min mogelijk op de voorstelling zelf letten.
Dus:
10.0 = uitstekend (maximaal)
0.1 = erbarmelijk (minimaal)

N.B.:
U mag dus een cijfer achter de komma geven.

In totaal krijgt U 118 dia's te zien, gedurende 15 sekonden elk.
De eerste zes dia's zijn proefdia's, en zij geven een een goed beeld van de kwaliteits-variatie van de resterende 112 dia's.

Als U nog vragen hebt, stel die dan gerust aan de proefleider. Anders, veel succes!

## B Lens defocusing

### B.1 Theory

Geometric optics give for a lens a description of the different types of deviation from an ideal image: the Seidel aberrations (see for example Born and Wolf (1 ). These are dependent on two variables: the position $\vec{r}$ of the point depicted on the screen and the place where the light beam passes through the lens surface $\vec{s}$ (both $\vec{r}$ and $\vec{s}$ are vectors). Röhler (8) describes in addition the consequences of defocusing of the lens. He gives for the total aberration function $\Delta(\vec{s},\vec{r})$ a series development:

$$\Delta(\vec{s},\vec{r}) = a_0+b_0r^2+b_1s^2+b_2(\vec{s}.\vec{r})+c_0r^4+c_1s^4+c_2(\vec{s}.\vec{r})^2+ c_3s^2r^2+c_4r^2(\vec{s}.\vec{r})+c_5s^2(\vec{s}.\vec{r})+d_0r^6+d_1s^6+\ldots$$

He also uses this to calculate the MTF of the system as follows:

$$\mathrm{MTF}(\omega) = \int_{\text{lens surface}} \exp(-2\pi i\,.\omega/\lambda\,.\,\delta\Delta\ (S_x,S_y,r)/\delta s_x\ ds_x\ ds_y$$

where $\omega$ is the spatial frequency on the image and $\lambda$ the wavelength of the light. From these two formulae it follows that first and foremost only the coefficients $b_1$, $c_1$, $c_2$, $c_3$ and $c_5$ influence the MTF. They stand for the effect as a result of defocusing and for the Seidel coefficients for spherical aberration, for astigmatism, for bending of the field and for coma, respectively. Röhler calculates further that the defocusing coefficient $b_1$ is proportional to the shift of the object plane from the ideal position $\Delta x$:

$$b_1 = B.\Delta x,$$

where B is a constant.

By Fourier transformation of the MTF the point spread function of the system can be calculated. On the basis of the above formula for the MTF we anticipate for the width $\sigma$ of the point spread function influences of the following effects:

- Defocussing gives a $\sigma$ which is proportional to the shift of the subject plane $\Delta x$.

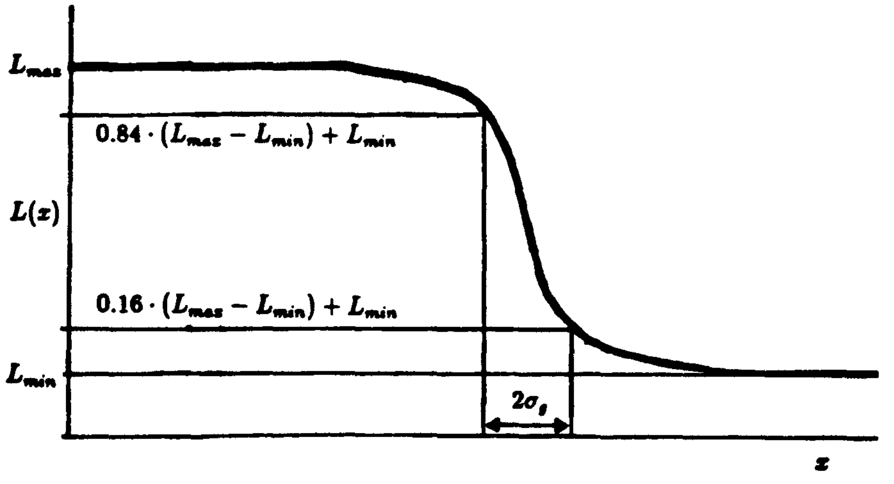



Figure 19: Cumulative Gaussian distribution
The cumulative Gaussian distribution for the luminance is described by $\int_{-\infty}^{x} P(x')dx'$, accent in which P(x) represents the Gaussian point spread function (see text). It is also indicated between what fractions of the luminance difference a width of $\sigma_g$ is calculated.

- Spherical aberration gives the same value for $\sigma$ over the whole image plane.

- Astigmatism etc. give rise to the dependence of $\sigma$ on the position of the image plane.

## B.2 Measurement procedure

For determining the MTF of the lens system the following method is applied.

A slide with a shaver blade in it (black-white transition) is projected onto the screen, by which a stepwise response is created. Perpendicular to the transition a Pritchard 1980 A-PL luminance meter, equipped with an extra SL-10A lens records the luminance. For this purpose the luminance meter is positioned on an x-y table which, computer controlled via a stepping motor carries out small discrete displacements of the order of 0.01 mm, and parallel to the screen (x direction).

In this way the step response S(x) of the projection system is measured. From this the 6 dB cut-off frequency $f_{6dB}$ (~/m) of the respective MTF is determined in two ways:

- If we assume that the point spread function of the system is a Gaussian distribution with spread $\sigma_g$:

$$P(x) = e^{-x^2/\sigma_g^2},$$

  then the stepwise response is described by a cumulative Gaussian distribution (see figure 19). The distance between the points where the luminance has risen by 16% and 84% of the maximum luminance difference, is equal to the root twice the width of the assumed Gaussian point spread function: $\sqrt{2}\sigma_g$. Still under the assumption of a Gaussian point spread function, the 6 dB cut-off frequency can be calculated from this value for $\sigma_g$ via:

$$f_{6dB,g} = \sqrt{\ln 2}\,/\pi\sigma_g$$

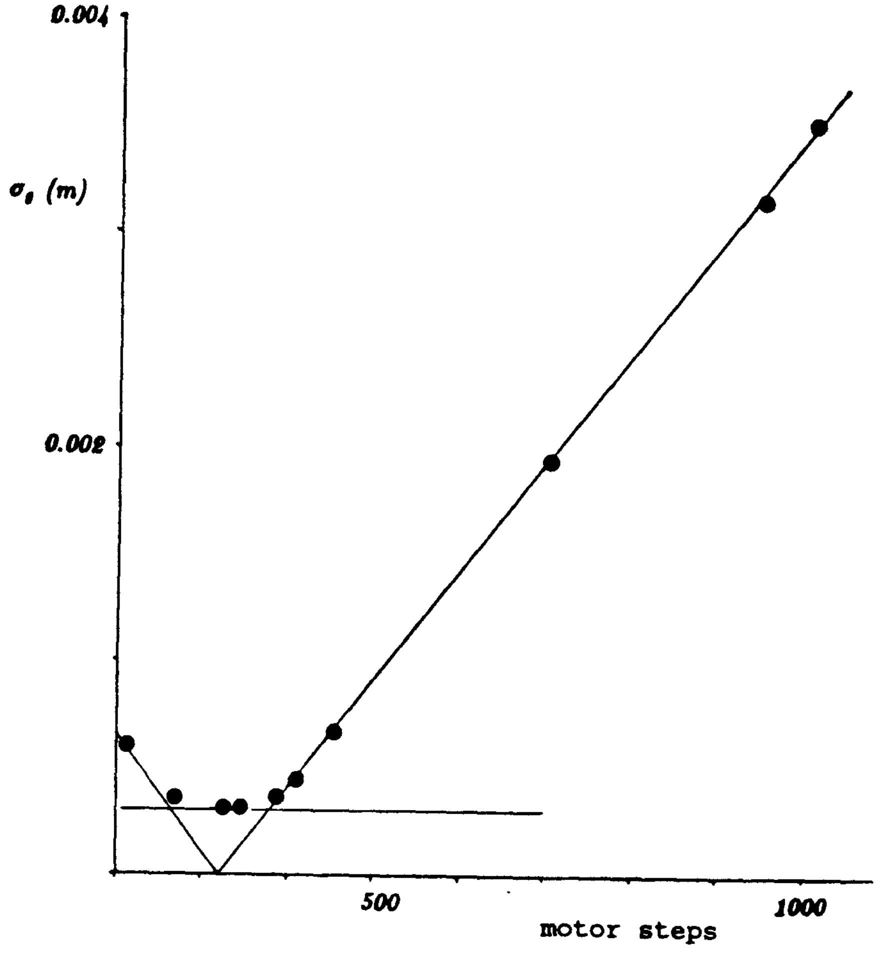



Figure 20: $\sigma_g$ as a function of defocusing
The defocusing is expressed in the number of motor steps, whereby 256 motor steps correspond to 1 mm. The lens in question is thus defocused at motor step 320. The horizontal and oblique asymptotes indicate the resolving power of the lens and the decreasing resolution as a consequence of defocusing, respectively.

- It is also possible to calculate the MTF of the system from the measured stepwise response via differentiation and Fourier transformation. From this the 6 dB cut-off frequency $f_{6dB,m}$ can be easily deduced.

### B.3 Measured resolutions and sources of spread

The 6 dB cut-off frequencies were determined for various defocusings and for various places on the screen.

Figure 20 shows the course of $\sigma_g$ with the defocusing. The data are described as anticipated by two asymptotes, determined by the resolving power of the lens (horizontal, see paragraph B.1) and by the defocusing (oblique) respectively. In the transition area between the two asymptotes it applies that $f_{6dB,g}$ and $f_{6dB,m}$ are not always equal, which is to be explained by the fact that the MTF measured is not Gaussian. In the other cases $f_{6dB,g}$ and $f_{6dB,m}$ agree very well, which indicates a Gaussian MTF.

In general it is also found that $\sigma_g$ is barely dependent on the position in the plane of projection. Measured in the radial direction (from the centre outwards) at the edge of the slide, some effects occur (coma) in the case of focused lenses. However, these only increase the value of $\sigma_g$ by a maximum of 20% (which even for a good lens appears to be normal), and disappear on defocusing.
The spread found in $\sigma_g$ in the case of replications can be described by the following sources:

- The spread due to the measurement procedure is approximately 1%. This value is caused by vibrations of the luminance meter, unevenesses in the projection screen and the like.

- The spread as a result of the continuous repositioning of the lens is dependent on the position of the lens: it varies from 5% in focused position to 1% in the case of strong defocusing. This spread is not a consequence of the positioning of the x-y table on which the lens is mounted, which is in fact done very accurately. The cause is more likely to be found in the vibration of the lens during positioning, which can result in slight differences in the ultimate position in relation to the slide.

- The changing of the slides gives a spread in the position where the oblique asymptote from figure 20 cuts the x axis (optimal focusing). This results in a defocusing-dependent spread in $\sigma_g$, which in percentage terms is maximum in the transition area between the two asymptotes (approximately 7%).

- Small and unavoidable jolts against the projector or the projection screen also result in an uncertainty in $\sigma_g$. For this reason all lens positions were calibrated both before and after the experiment. The difference between the two calibrations is however in many cases not to be explained in a systematic manner. In practice, the spread measured in this way appears to be the most significant.