Perceptually weighted spatial resolution

Instituut voor Perceptie Onderzoek
Postbus 513 - 5600 MB Eindhoven

JW/jw 88/10
22.01.1988

Rapport 587

Perceptually weighted
spatial resolution

J.H.D.M. Westerink

Rapport no. 587

Perceptually weighted spatial
resolution

J.H.D.M. Westerink

SUMMARY

This report is the result of a literature study into measures of resolution of imaging equipment which correlate with the subjective sharpness sensation. A large number of these resolution measures are described, along with a number of studies which compare various resolution measures on this point.

It emerges that in general a high correlation can be attained between measure of resolution and subjective sharpness. This applies in particular to the MTFA measure for which correlations of between 0.84 and 0.92 are reported.

However, a number of reservations are made in respect of these conclusions and attention is also drawn to a number of as yet unresolved problems.

CONTENTS

1 Introduction

Of the many factors which play a role in picture quality, resolution is one of the most important. There is a wide range of formulas for expressing resolution and all are to a greater or lesser extent suitable for the environment in which they are used. This report takes stock of measures for the resolution of imaging systems, such as picture screens, photo-printing processes and projection systems. There are two limiting conditions:

- The resolution measure applies in the first instance to the imaging system itself and not the scene depicted. The intention is to translate the physical parameters of the imaging system into a measure of resolution which is independent of the (continually changing) scene content.

- The resolution measure must be perceptually relevant. This means that it must correlate with the subjective sharpness which is experienced by the observer.

Over the course of the years many resolution measures have been developed specifically for imaging systems. They emerged either from the photographic industry or from the electro-optical industry. Now that the Modulation Transfer Function (MTF) has gained in familiarity and popularity in both areas, there has however been a cross-fertilisation of ideas.

The perceptual evaluation of the different resolution measures and their comparison has also been approached from two different angles. First and foremost this was done in military circles, where great emphasis is placed on detection and discrimination (task-oriented environment). It is not surprising that in this environment perceptual evaluation too is often based on identification tasks. It is a completely different case in the consumer-oriented industry, where no direct performance of any nature can be linked to the normal use of the imaging system (non-performance environment). This means that evaluations are more often than not based on the judgments of test subjects. The division is not however totally clear-cut: in military circles use is still made occasionally of the judgments of test subjects, while in non-task-oriented environments there is also some evaluation on the basis of discrimination experiments.

The perceptual attribute corresponding to the physical dimension of "resolution" we refer to as "sharpness", and this forms part of the more complex psychological concept of "quality". The sensation of sharpness can however be influenced by physical parameters such as luminance or contrast. But these parameters also directly influence the quality, for example via the psychological concept of "brightness". The literature often fails to make a clear distinction between the concepts of quality and sharpness, which in certain situations can lead to misunderstandings. On the other hand, this is the reason why resolution measures which correlate with the broader concept of quality are taken into consideration.

Whereas the perceptual attributes are restricted to sharpness and quality, the physical characterisation is based on a broad scale of parameters. These are:

- x',y': spatial coordinates of the scene [m]. In this report the word scene relates to a three-dimensional reality. From a certain standpoint this can however be described by means of a two-dimensional projection. The concept of scene is in contrast to the concept of image, which is the (end) result of the whole imaging system.

- u',v',w': spatial frequencies[m^{-1}], corresponding to the coordinates x' and y'. The parameter w' (often) describes a radial frequency.

- L'(x',y') : luminance distribution of the scene [cd/m^2].

- L'_{max},L'_{min} : maximum and minimum luminance in the scene [cd/m^2].

- m_0 : modulation or contrast modulation of the scene. This is calculated as $(L'_{max}-L'_{min}) / (L'_{max} + L'_{min})$.

- S(u',v'),S(w') : Fourier transform of L'(x',y'). S(u',v') denotes the spatial frequency content of the scene.

- d : width of image [m].

- x,y,r : coordinates of the image [m]. These are linked to the coordinates of the scene, x',y' and r', and vice versa. In this case too, r (often) denotes a radial parameter.

- u, v, w : spatial frequencies of the image $[m^{-1}]$, corresponding to x, y and r. These are linked to the spatial frequencies of the scene, u', v' and w'.

- L : average luminance of the image $[cd/m^2]$.

- γ : gamma of the imaging system. This gives the relationship between the luminance of the scene and that of the image: $L \propto L'^{\gamma}$. This relationship can be used for a large range of luminances for most imaging systems. In view of the fact that γ is usually greater than 1, it is clear that these imaging systems are not linear.

- $j(r), j(x,y)$: profile of the spot of a CRT.

- $PSF_s(x,y), PSF_s(r)$: point spread function of the imaging system.

- $MTF_s(u,v), MTF_s(w)$: modulation transfer function of the imaging system. This MTF can be calculated as the Fourier transform of the point spread function of the imaging system. It is usually normalised so that $MTF_s(0)=1$. It is not always clear a priori that the MTF exists: for beneficial use of the MTF it is necessary for the system to be linear. This MTF must also be position-independent - i.e. homgeneous and isotropic - , if it is to describe the whole image. In the case of non-linear systems the MTF is often used as a first-order approximation. The great advantage of working with descriptions on MTF basis is that by multiplying the MTFs of the various system components it is possible to obtain the MTF of the whole system.

- MTF_{s1}, MTF_{s2}, MTF_{si}: MTFs of the various system components.

- $N_s(u,v), N_s(w)$: Wiener noise spectrum of the imaging system.

- a : viewing distance $[m]$.

- $m_{d,v+s}(u,v), m_{d,v+s}(w)$: minimum contrast modulation required for detection of a sine raster with spatial frequency w, or u and v

(two-dimensional). This threshold value is measured at a fixed, generally optimised viewing distance at the output of the imaging system and is therefore determined both by characteristics of the imaging system (for example noise, light intensity) and by the visual system. The index d in this context relates to the fact that it concerns a threshold value, while the indices v+s indicate that in the determination of this threshold both the visual system (v) and the imaging system (s) play a role.

- μ,ν,ω : spatial frequencies in the eye [per/degree] (or possibly [per/mm on the retina]). At a known viewing distance these can be converted into the spatial frequencies of the image according to the formula $u=360\cdot\mu/2\pi a$, and in the same way for v and ν, and w and ω.

- $m_{d,v1}(L,\mu,\nu)$, $m_{d,v1}(L,\omega)$: contrast modulation threshold for detection of sine rasters. This value is of course also dependent on the average luminance L. The reciprocal value of the contrast modulation threshold is (whether normalised or not) considered as the contrast sensitivity of the eye.

- $m_{d,v2}(L,\mu,\nu)$, $m_{d,v2}(L,\omega)$: contrast modulation threshold for discrimination between a sine and a block raster of the same frequency.

- $C_{v1}(L,\mu,\nu)$, $C_{v1}(L,\omega)$: contrast sensitivity of the eye. This is in fact the reciprocal of the contrast modulation threshold $m_{d,v1}$, with an arbitrary normalisation dependent on the author.

2 Measures of resolution

This chapter considers a number of measures of resolution known in the literature. A division is made into two groups. The first includes a number of measures which take no account of the influence of the human eye on the perceived resolution (sharpness). These resolution measures are sometimes the direct result of a measurement procedure. In that case the resolution measures in question are not usually based on MTFs. However, when it became possible and common practice to measure the MTF of a system, there arose the problem of the interpretation of this wealth of data. A large group of measures of resolution was accordingly introduced in an attempt to sum up the MTF data in a single figure which was to indicate the quality or sharpness of the imaging system.

Once the concept of MTF was accepted in the literature as a possible way of describing the resolution, it was also realised that one important factor in the imaging system had so far been neglected, namely the eye of the person looking at the image. The second group of resolution measures is characterised by the fact that they try to incorporate data on the visual system in the construction of resolution measures. A number of extra parameters are thereby introduced, such as the luminance **L** and the viewing distance **a**, which must provide for the link between the visual and the imaging system. It then became generally accepted, due to the introduction of the visual system into the resolution measure, that the suitability of the proposed measure is dependent on its correlation with subjective sharpness or quality.

Both types are considered separately in a compilation of resolution measures in the following two sections.

2.1 Measures based solely on physical measurements

The physical measurement for the determination of resolution can be done in several ways. One of the most popular methods is to establish the MTF of the system. The use of the MTF has certain advantages, the most important being that it gives a description of the system in orthogonal base functions: sines. The total MTF of a number of systems arranged in succession can therefore be easily calculated by multiplying the individual MTFs.

Balanced against these advantages there is however a number of disadvantages. An important requirement for the use of MTFs is that the system in question should be linear, homogeneous and isotropic. In photography and electronic image reproduction this condition is rarely fulfilled. The description of such imaging systems in terms of MTFs is accordingly often used as a first-order approximation.

The fact that the MTF is an overall measure can also be considered as a disadvantage: in a direct measurement of the MTF the effects on the various positions on the image (inhomogeneity, anisotropy) are averaged out to an overall value which does not therefore correspond to a specific place on the image. A direct measurement of the MTF of an imaging system also becomes inaccurate for the lower spatial frequencies due to the window effect of the size of the image.

A description in the form of a point spread function (PSF) is a possible alternative to the MTF. The main advantage of the use of the PSF is that it gives a local description. This dispenses with the conditions that the imaging system must be homogeneous and isotropic, because "only" a local description of the system is given. The linearity of the system however remains a prerequisite. The PSF of a number of systems connected in succession can be calculated by the convolution of the individual PSFs. The execution of this convolution is of course quite feasible, but somewhat more laborious than the multipication of a number of MTFs. This is often considered as a disadvantage of a description in terms of PSFs. In the measurement of the PSF the size of the image may also exert a detrimental window effect, as is the case with the measurement of the MTF.

If it emerges after measurement of the PSF that the system is homogeneous and isotropic, then the MTF of that system can be determined through a Fourier transformation from the PSF. In that case the two descriptions are therefore equivalent. It is more often than not the case that the MTF which is indicated for an imaging system is calculated in this way from a measurement of the PSF.

In the following summary of measures of resolution it will emerge that by far the most are based on a description based on the MTF. This is despite the fact that the MTF and PSF description methods are very closely related, and despite the fact that the imaging equipment in question does not usually fulfil the conditions for the use of the MTF.

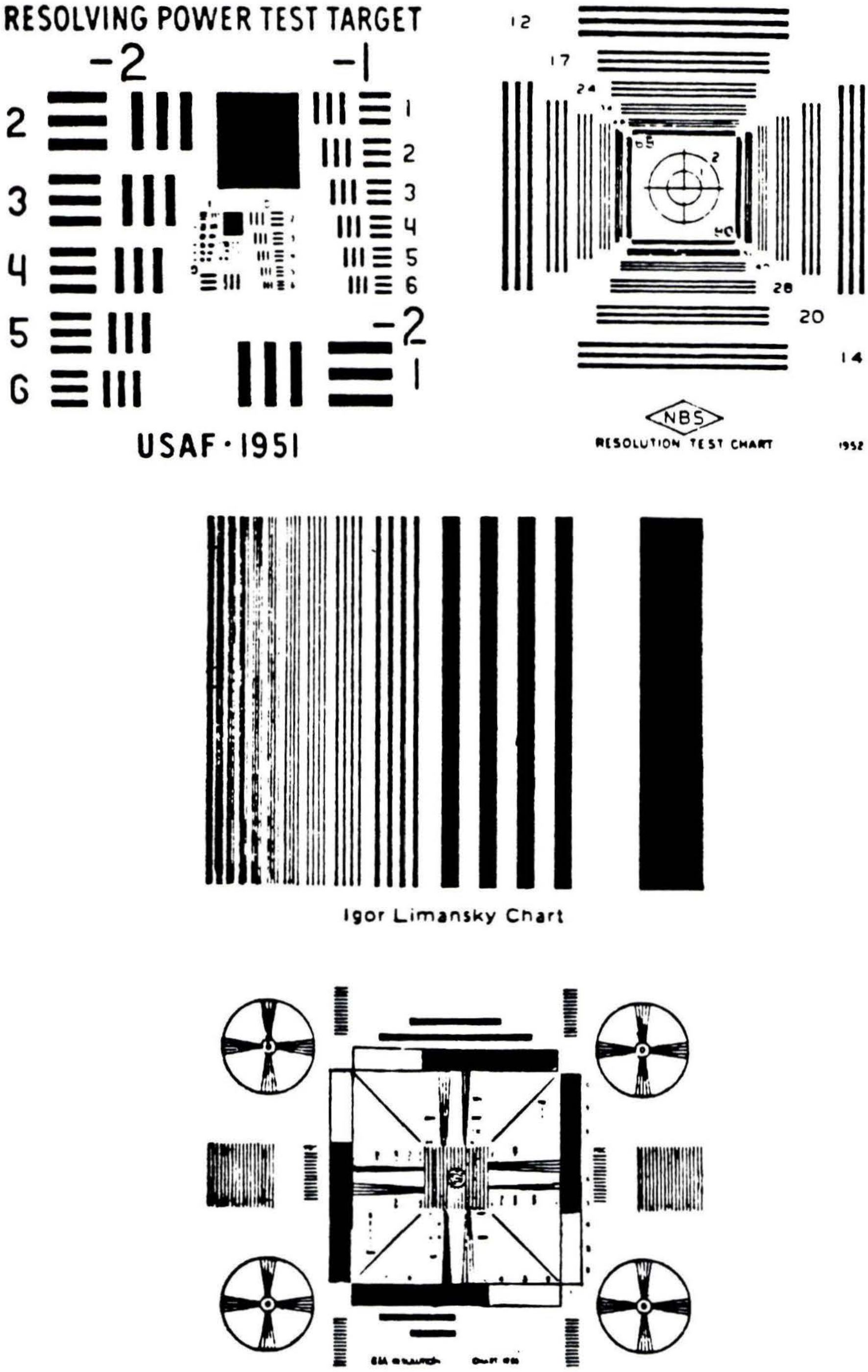

Figure 1 : A selection of resolution charts

2.1.1 Maximum line density

Both in the photographic industry and in the world of picture tubes it has long been common practice to express resolution in the maximum number of lines to be reproduced per unit of length. These measures are still encountered in specifications today. The value of this maximum line density is determined on the basis of test patterns. A selection of test charts is shown in figure 1.

2.1.2 Spot width

The spot width is essentially used in characterising the picture tubes. This measure indicates the width of the spot, either determined by the eye or measured with equipment which scans the spot profile j(x,y) at a certain percentage of the maximum profile height. It is commonly the half profile height which is used, but there are also many other percentages in circulation. Barten (1) proposes

$$d_{0.05},$$

the width of the profile, measured at 5% of the maximum profile height. Figure 2 sketches a number of different profiles which all have the same $d_{0.05}$. If we assume that the spot profile is the sole determining factor for the MTF of the picture tube, then it can be considered as a point spread function and the MTF can be calculated by Fourier transformation. The MTFs calculated in this way for the profiles in figure 2 are almost identical for low frequencies, which is not surprising. The profiles coincide at a low value (0.05) and therefore at a large width. Because the lower spatial frequencies correspond with the larger wavelengths, these identical profile values at large widths result in an identical curve of the MTF at low frequencies (provided of course that any bizarre profile shapes are disregarded). In mathematical terms the MTF values for the low spatial frequencies are to be estimated as follows:

$$\mathrm{MTF}(w) = 1 - 2\pi s^2 w^2,$$

where

$$s^2 = \frac{\int_0^\infty 2\pi r^3 j(r)dr}{\int_0^\infty 2\pi r j(r)dr} .$$

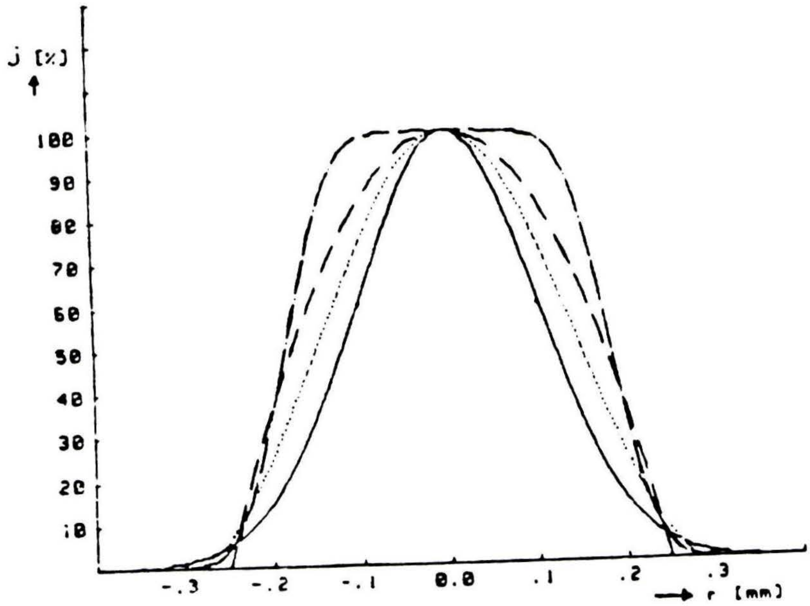

Figure 2: Various spot profiles with the same $d_{0\cdot 05}$

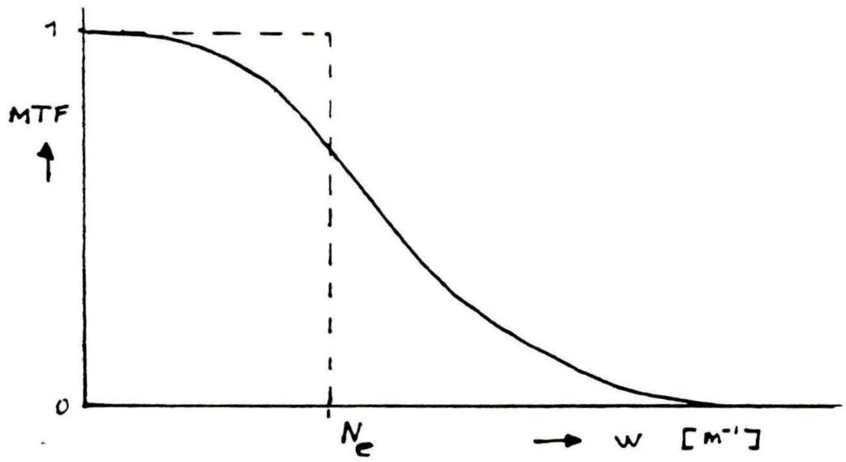

Figure 3: MTF and noise equivalent passband

In fact s^2 is the third-order moment of the Hankel transformation of the profile.

In the resolution measure described, $d_{0.05}$, the low spatial frequencies therefore play a more important role than the higher spatial frequencies. Barten claims that this is also the case in the visual system and states moreover that $d_{0.05}$ is also precisely that width which is perceived with the eye.

2.1.3. Noise equivalent passband

At the beginning of the fifties Schade published a series of articles (25) (26) in which he introduced the use of the MTF for the qualification of imaging systems. He then based a summarising measure of resolution on the MTF, the "noise equivalent passband":

$$N_e = \int_0^\infty MTF_s^2(w)dw.$$

The formula calculates the cut-off frequency of an (idealised) rectangular MTF, which gives the same power as MTF_s, see figure 3. Contrary to what the name suggests, N_e does not describe any effects of noise in the system.

To take account of various system components two methods can be used, which according to Schade give the same value within 5% for the noise equivalent passband. Either the MTF of the system $MTF_s(w)$ is calculated from the $MTF_{si}(w)$ of the various subsystems via multiplication, and from this the value of $N_{e,s}$. Or a separate $N_{e,si}$ is calculated for each system component and these are then added up according to

$$N_{e,s}^{-2} = N_{e,s_1}^{-2} + N_{e,s_2}^{-2} + \ldots + N_{e,s_i}^{-2}.$$

By assigning the visual system its own noise equivalent passband Schade has also found a possibility of including the influence of the eye in the measure of resolution. He states that for luminances between 15 and 35 cd/m^2:

$$N_{e,eye} = 752.d/a,$$

applies, where d denotes the image size and a the viewing distance. This formula for the noise equivalent passband of the eye $N_{e,eye}$ is not however dimensionally accurate, unless it is assumed that the constant 752 has the dimension of a spatial frequency. Schade includes $N_{e,eye}$ as an extra system component, equivalent to all other system components. Introduced in this manner, the eye is described as a standard against which the resolution of the image is measured and which in this way introduces a type of saturation. There is however no question of a weighting of the system MTF with the visual frequency response curves. Nor does Schade have any method for arriving at any sort of optimal viewing distance: given the noise equivalent passbands of the various subsystems, the maximal noise equivalent passband of the system is always found for a viewing distance of 0 m.

2.1.4 Equivalent width

Bracewell (5) describes a number of measures which in some way or other constitute a measure for the width of a point spread function or its Fourier transform. The most well-known of these is the equivalent width:

$$W_e = \frac{\int_{-\infty}^{\infty} PSF_s(r)dr}{PSF_s(0)} = \frac{MTF_s(0)}{\int_{-\infty}^{\infty} MTF_s(w)dw},$$

which can be calculated both from the point spread function and from its Fourier transform. Another measure proposed by him is the mean-square width:

$$W_{ms} = \sqrt{\frac{\int_{-\infty}^{\infty} PSF_s(r)^2 dr}{\int_{-\infty}^{\infty} PSF_s(r)dr}},$$

which in the same way as the $d_{0.05}$ of Barten (paragraph 2.1.2) places more emphasis on the lower spatial frequencies than on the higher. Bracewell did not however intend the aforementioned measures specifically for imaging equipment, but simply as a general measure for the width of a function or its Fourier transform. In principle he could also have suggested the reverse on the basis of the MTF, for example:

$$F_{ms} = \frac{\int_{-\infty}^{\infty} MTF_s(w)w^2 dw}{\int_{-\infty}^{\infty} MTF_s(w)dw}.$$

Bracewell himself makes no use of these measures when he describes the applications of Fourier transformations in TV technology in his book.

2.1.5 Information fidelity

Linfoot (20) describes a number of measures which indicate the quality of the imaging system. Because he concentrates in his book on Fourier methods, he describes his measures of resolution both in terms of spatial coordinates and in terms of spatial frequencies. The accuracy of the image, information fidelity, is calculated as follows:

$$IF = 1 - \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |1 - MTF_S(u,v)|^2 |S(u,v)|^2 dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |S(u,v)|^2 dudv} .$$

The term with the ratio of integrals is called the fidelity defect. The squared picture fault is calculated as the square of the difference between the system MTF and an ideal MTF (always 1), weighted according to the spectral content of the scene S(u,v), and integrated over all spatial frequencies. The integral is also normalised to the influence of the weighting factor. In this form the resolution measure does not meet the requirement that it can be calculated independently of the scene. With a small modification however, the information fidelity can be rendered suitable as a measure of resolution for describing the system. One possibility is for the present scene contents of S(u,v) to be replaced by an average of all conceivable scene contents, as for example Carlson (7) does in later work (see paragraph 2.2.8).

In addition to the information fidelity Linfoot also introduces the concept of structure content, which should provide a measure for the structure present in the image, relative to the structure of the original scene:

$$SC = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |MTF_S(u,v)|^2 dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |S(u,v)|^2 dudv} .$$

An interpretation of this structure content is not however given by Linfoot. The formula is introduced essentially on mathematical

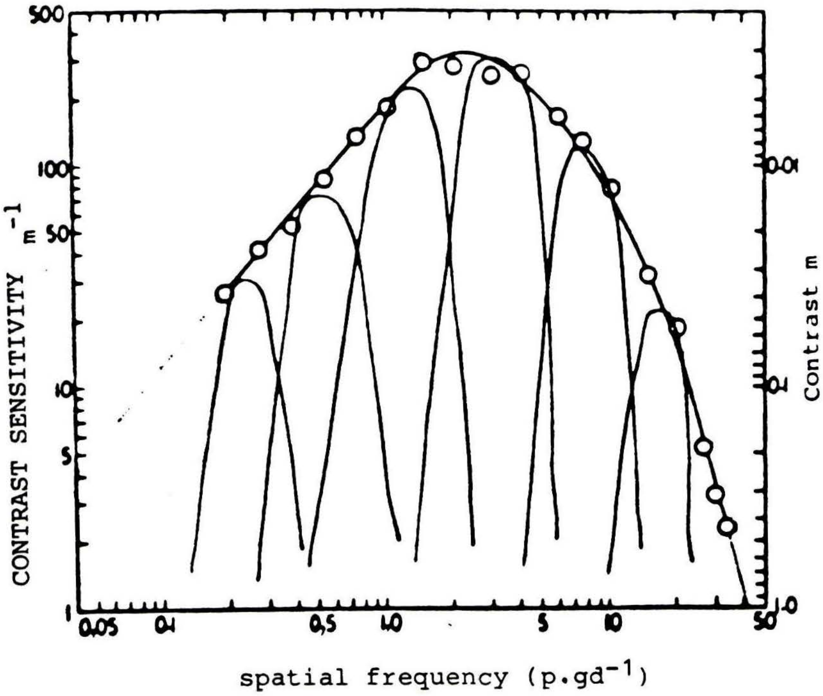

Figure 4: Spatial channels in the visual system

The measurement points indicated represent the contrast sensitivity of the visual system. A number of frequency-specific band filters are drawn in. They are referred to as channels and are together responsible for the contrast sensitivity measured.

grounds. This also applies to his third proposal, the correlation quality, in which the measure of correlation between image and original scene is calculated:

$$CQ = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |MTF_S(u,v)S(u,v)|^2 dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |S(u,v)|^2 dudv} .$$

The three measures have a mutual relationship according to the formula:

$$IF + SC = 2CQ.$$

2.2 Measures based on the visual system

A number of resolution measures is based on knowledge of the functioning of the eye. An initial inspection of our habitual manner of viewing reveals that we tend to focus our attention on a specific point in the picture, while we can at the same time take in some overall structure. The eye is therefore capable of providing us with both local and global information at the same time. It can be deduced from this that the visual system comprises a number of processors, which allow a certain latitude both in the spatial domain and in the domain of the spatial frequencies.

At the moment the spatial part of the visual system is in general consensus described as being built up from a number of receptive fields in the eye, with different retina positions and different spatial resonance frequencies. It is therefore assumed that the eye functions by means of a number of bandpass filters operating in parallel, called channels (see figure 4). These channels together determine the contrast sensitivity, which is determined as a function of the spatial frequency and is defined as the reciprocal of the contrast modulation threshold for a sinusoidal raster of that frequency. The contrast sensitivity is also dependent on the average luminance level and generally decreases for very high and very low spatial frequencies.

The contrast sensitivity - normalised or not - is often considered as the MTF of the eye. This is incorrect for various reasons. First of all, the eye is in that case described as one low-pass filter, whereas in reality it consists of a number of bandpass filters. It is also the case that the contrast sensitivity is determined from measurements at threshold level: it is not clear a priori that the various channels work together in the same way at suprathreshold level (where our interest lies). On the contrary, comparative experiments by, for example, Watanabe (34) have shown that at suprathreshold modulations the "suprathreshold contrast sensitivity" tends increasingly to assume a low-pass character. Thirdly, the eye is largely inhomogeneous (see for example Davson (13)), the reason why, in the same way as for image reproduction equipment (section 2.1), a more local description is required.

Another method for describing the visual system is by means of a point spread function. This can be measured directly with the aid of a perturbation technique and using points and lines as stimuli (see for example Blommaert (4)). Due to the spatial frequency content of these stimuli, only the PSF of the narrowest, or the most high-frequency channel is then determined. One disadvantage of this is that the point spread function thus as a whole gives no information on the functioning of the visual system at the lower spatial frequencies. On the other hand, it could be assumed that for the sharpness percept it is precisely this narrowest channel which is important.

A third possibility for characterising the response of the visual system is by means of Gabor functions. A Gabor function consists of a sinusoidal spatial frequency which is modulated by a Gaussian envelope of which the width is proportionate to the wavelength of the sine. The Gabor function therefore has, in the same way as the visual system, an extensiveness in both the spatial domain and in that of the spatial frequencies. It is for this reason that the Gabor function is proposed and studied as a basic unit for visual perception (see for example Watson (35)).

It is clear that the optimal perceptually relevant resolution measure must be based on a description of the imaging system which ties in with the manner in which the visual system processes the information. For this it is important to know what the basic functions of the visual system are, but at the moment knowledge is still insufficient. In practice there appears to be more of a reverse dependency: precisely because linear systems can be so well described in the frequency domain, attempts have been made to also

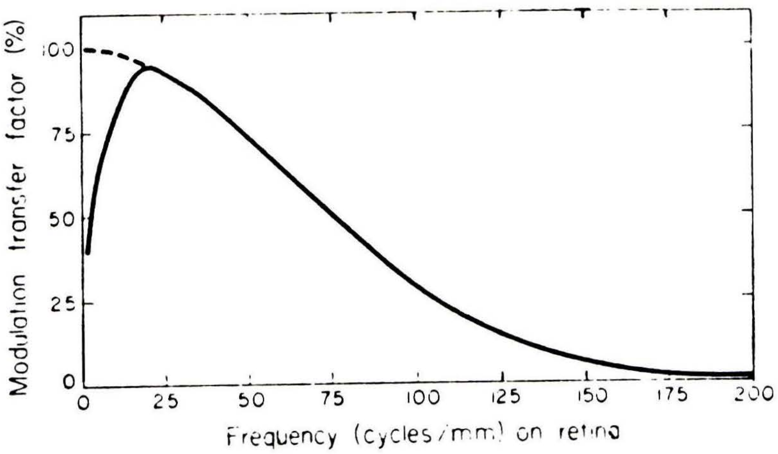

Figure 5: Standard curve for MTF_s observer (w).
This curve is presented on the basis of contrast sensitivity measurements by Schade, Lowry and DePalma, and Wolfe (drawn line). Due to "compensating intellectual processes" the curve is modified for low spatial frequencies (dashed line). The curves are considered to be applicable for luminances between about 50 and 150 cd/m^2.

describe the visual system on the basis of sine rasters. This will also emerge in the resolution measures which are considered in the following sections: the visual system is almost always described by means of the contrast sensitivity and, what is more, this is often interpreted as an MTF of an extra imaging system.

2.2.1 SMT and CMT acutance

One of the first resolution measures of which it was claimed that it correlated with (subjective) sharpness was the "system modulation transfer acutance" (SMT acutance) of Crane (11). This was used a great deal in the photographic industry, although it was probably based more on practical experience than on system analyses. It is calculated as follows from the MTFs of the various subsystems:

$$SMT = 120 - 25\log\left(\sum_{i=camera}^{observer}\left(\frac{200.N_i}{\int_o^{\infty} MTF_{si}(w)dw}\right)^2\right)$$

N_i is in this formula a magnification factor representing the relationship between the picture width on the retina and that in system component i. The magnification factor ensures that the MTFs of the various system components are now expressed in spatial frequencies on the retina.

The $MTF_{s,observer}(w)$, which Crane uses is an "optimistic compromise" between measurements of contrast sensitivity by Schade (26), Lowry and DePalma (21) and Wolfe (36). On account of assumed "intellectual processes" which are said to compensate for the poor sensitivity at the lower spatial frequencies, a substantial modification is applied in that area, as is shown by figure 5. It is assumed that the curve is largely applicable for luminances between 50 and 150 cd/m^2.

The way in which the influences of the various system components are summed up (N.B.: the logarithm is taken over the complete sum) is similar to the method used by Schade (see section 2.1.3.). In the same way as Schade, Crane includes the influence of the visual system in a final separate term in the summation. Here too, this visual term only has a saturating influence, so there is no question of any

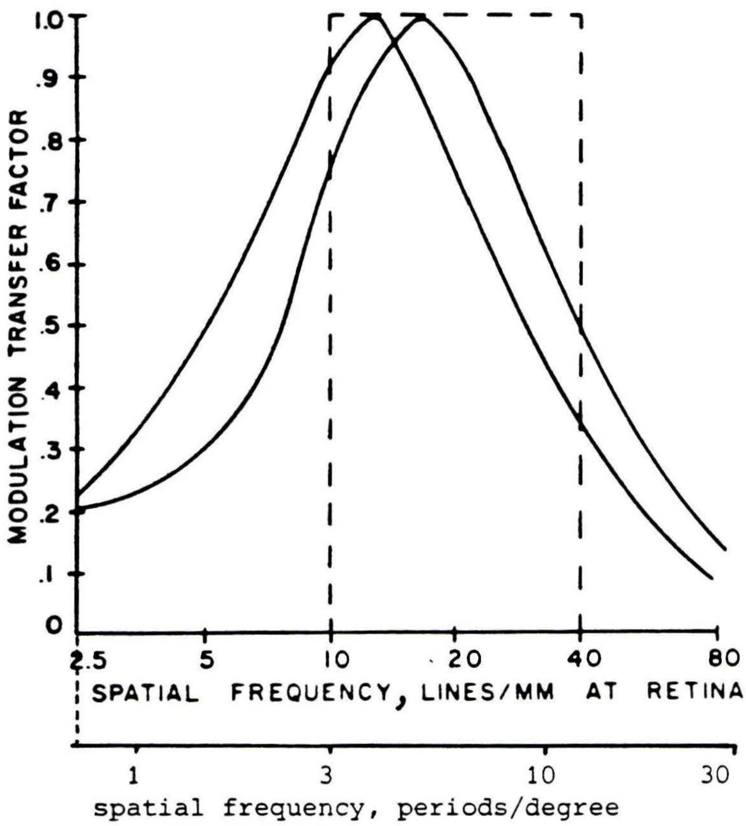

Figure 6: Contrast sensitivity curves
They are measured at 15 (left-hand curve) and 150 cd/m^2. Within the dashed lines is the frequency zone which is included in the SQF. The spatial frequency is indicated in both lines/mm at retina and periods/degree.

real influence of visual data when the MTF of the imaging system is changed. Contrary to Schade, however, Crane places the emphasis on the correlation with subjective sharpness.

Some years later Gendron (14) presented a related measure of resolution, the "cascaded modulation transfer acutance" (CMT acutance)

$$CMT = 125 - 20 \log \left(\frac{200}{\int_0^{\infty} MTF_s(w)dw} \right)^2,$$

where $MTF_s(w)$ is the product of the MTFs of the various system portions, including $MTF_{observer}(w)$. The upper limit of the integration path, here ∞ , can according to Gendron also be determined by the visual system. Gendron claims for his CMT acutance a higher correlation with subjective sharpness judgments than is provided by the SMT acutance. This is primarily due to the correct processing of system MTFs with an "overshoot" (values greater than 1) which in general do not correlate well with the SMT acutance. The comparative experiment that led to this conclusion is however only briefly described.

2.2.2 Subjective quality factor

Granger and Cupery (15) introduce a measure which solely by its integration limits recalls the visual system, as is shown by figure 6. The upper and lower limit are set, on the basis of measurements by Schade (27) on the contrast sensitivity of the eye, at 10 and 40 periods/mm on the retina. At a distance of the retina from the lens focus of 17 mm, these values correspond with spatial frequencies of 3 and 12 periods/degree respectively. At a known viewing distance these values can of course in turn be converted into spatial frequencies w_3 and w_4 of the image. The integration over the spatial frequencies is on logarithmic basis, according to the hypothesis that Weber's law is also applicable to an aspect not specified further, relating to the spatial frequency axis. In this way the subjective quality factor is found:

$$SQF = K \int_{w_3}^{w_4} \int_0^{2\pi} |MTF_s(w, \theta)| d\theta d(\log w),$$

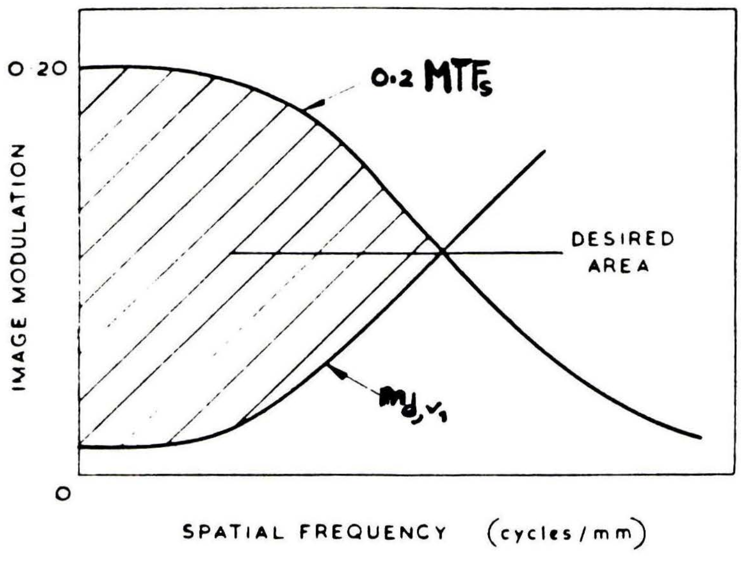

Figure 7: Calculation of the TQF

The TQF is in fact the area of the shaded zone.

where K is a normalisation constant. Possible effects of anisotropy are included by also integrating over the orientation angle θ.

Granger and Cupery have attempted to demonstrate the use of the SQF in an experiment based on pair comparisons. They find for the SQF a correlation of 0.988. It is however unclear with what the SQF then correlates, in view of the fact that they first refer to the sharpness and then to the quality again.
There is also muddled and indiscriminate use of both subjective quality scales and units of just noticeable difference.

2.2.3 MTFA, TQF and measures deduced from them

Charman and Olin (9) introduced in 1965 the most modified resolution measure in history, the threshold quality factor:

$$TQF = \int_0^\infty (m_0 . MTF_s(w) - m_{d,v+s}(w))dw.$$

where $m_{d,v+s}(w)$ is the minimum contrast modulation in the image required for detection at a spatial frequency w. This is not only dependent on the eye, but is also determined by the noise in the imaging system. The curve therefore contains influences of both the visual system, principally in the lower spatial frequencies, and of the noise of the imaging system in the higher spatial frequencies.
m_0 is the contrast modulation of the target, defined as
$m_0=(L'_{max} - L'_{min})/(L'_{max} + L'_{min})$. In real photos m_0 therefore actually varies with the scene content. For the application of aerial photography Charman and Olin set the value of m_0 at an average of 0.2.

Although the integral runs to ∞ according to the definition, it emerges from figure 7 that what is referred to is the area between the two curves. The TQF therefore stands for the measure of suprathresholdness of the contrast modulation in the picture.

Charman and Olin deduced a formula for $m_{d,v+s}(w)$ for photographic images on the basis of parameters such as grain size G and emulsion density D of the film and the quantity of light E to which the film is exposed:

$$m_{d,v+s}(w) = 0.0034.\left(\frac{dD}{d(\log E)}\right)^{-1}.(0.33 + G^2S^2w^2)^{\frac{1}{2}}.$$

S here denotes the minimum signal-to-noise ratio required for detection and Charman and Olin set this at a value of 4.5. For electro-optical systems there is no general analytical formula known for $m_{d,v+s}(w)$ and the curve must therefore be determined separately for each system.

Snyder (30) incorporates the idea behind the TQF in the concept of his MTFA (modulation transfer function area):

$$MTFA = \int_0^{w_1} \left(MTF_s(w) - \frac{m_{d,v+s}(w)}{m_0} \right) dw,$$

where w_1 is the spatial frequency at which the two curves cross (see figure 7). The MTFA differs in principle by a factor of m_0 from the TQF. Synder has thereby shifted the emphasis from a description in terms of contrast modulations to one in terms of the system MTF.

Synder also applied the MTFA, which had demonstrated its usefulness in photography, to electro-optical pictures, where electronic noise assumes the role of the grain structure. For high signal-to-noise ratios the influence of the noise on the detection curve $m_{d,v+s}(w)$ appeared to be very small, this being the reason why later authors, for example Task (33), replaced this curve by the threshold curve of the visual system for detection of a sinusoidal modulation $m_{d,v1}(L,\omega)$. For a fixed viewing distance a the spatial frequencies of the eye ω can be converted to those on the image w. The unknown factor m_0 is then for convenience set at 1, which results in a formula which is often quoted in the literature as the MTFA, and which we shall indicate as:

$$MTFA_{lit} = \int_0^{w_1} \left(MTF_s(w) - m_{d,v} \left(L, \frac{2\,a}{360} \cdot w\right) \right) dw.$$

This formula is rather bizarre because the difference is calculated between an MTF and a contrast modulation, which only works well because both quantities are dimensionless. Moreover, the assumption $m_0=1$ together with the normalisation of $MTF_s(0)=1$ implies that the imaging system is capable of producing a contrast modulation with the value 1.

A large number of modifications of the MTF is proposed in order to

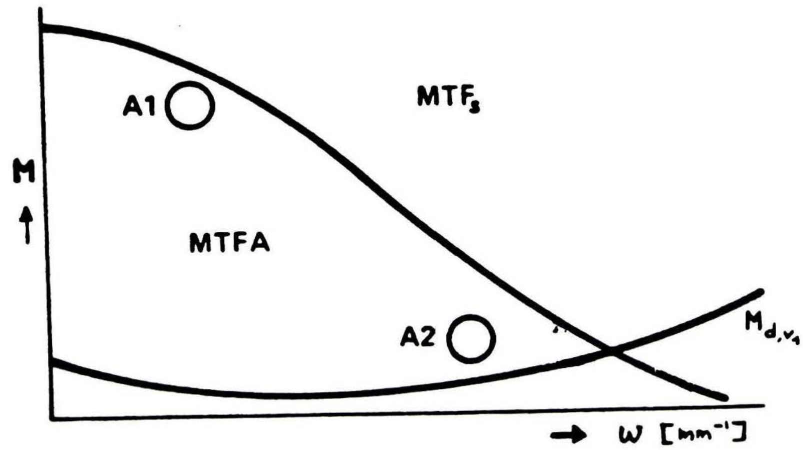

Figure 8: MTFA with modifications.

The zones A_1 and A_2 do not necessarily make the same contribution to the sharpness sensation, which can be expressed in the resolution measure.

regulate the extent to which the various spatial frequencies and the various levels of modulation contribute to the integral. The question is in what ratio the areas A_1 and A_2 in figure 8 "contribute" to the sharpness sensation and how this is to be incorporated in the measure of resolution. Synder (30) calculates, for example, a different surface area and therefore a different measure when the curves are plotted on logarithmic axes:

$$MTFA_{log-log} = \int_{\log w_0}^{\log w_1} \left(\log \frac{m_0 . MTF_s(w)}{m_{d,v+s}(w)} \right) d \log w,$$

where w_0 is needed as a lower limit and can for example be set at 10 periods/mm. In this formula the system modulation is actually divided by the threshold curve $m_{d,v+s}$ instead of this being subtracted. The resulting ratio is also the basis of a large number of other resolution measures, as will emerge from the following sections.

Task (33) even introduced three related measures. The "lower limit MTFA" is calculated in the same way as the MTFA, but the lower spatial frequencies are not considered to be important. For this reason integration takes place from a lower limit w_2, which is set at two periods/degree (and hence $w_2 = \omega_2 . 360/2\pi a$):

$$LLMTFA = \int_{w_2}^{w_1} \left((MTF_s(w) - m_{d,v_1}(\frac{2\pi a}{360} . w) \right) dw.$$

The 'band-limited MTFA' arises in the same way, but makes use of the threshold modulation curve for discrimation between a sine and a block raster $m_{d,v2}(\omega)$:

$$BLMTFA = \int_{w_2}^{w_1} \left((MTF_s(w) - m_{d,v_2}(\frac{2\pi a}{360} . w) \right) dw.$$

This BLMTFA is developed as a measure of resolution for task-oriented environments, in which discrimination and identification play a role. A good example is again military aerial photography: Task reasons that the perception of an object (tank) must be linked to a description in terms of threshold modulation curves for sine rasters. The identification of various types of tanks however requires more detail, or calls for a greater power of discrimination. According to Task this is therefore better described by the threshold modulation curve for discrimination between a sine and a block raster: the BLMTFA is to be considered as a measure for power of **detail discrimination.**

As the third measure Task presents the grey shade frequency product. Whereas in previous MTFA modifications it was always the influence of the frequency axis which was manipulated, in the GSFP an attempt is made to optimise the measure of resolution by actually adjusting the influence of the MTF axis:

$$GSFP = \int_0^{w_1} G(MTF_s(w))dw,$$

where

$$G(MTF_s(w)) = 1 + \frac{\log \frac{1+MTF_s(w)}{1-MTF_s(w)}}{\log \sqrt{2}} .$$

In this formula the MTF is again considered as a contrast modulation and this is subsequently converted to a grey shade range $G(MTF_S(w))$, in which it is assumed that one grey shade difference corresponds with a factor $\sqrt{2}$ in luminance. Making use of the definition of modulation $m = (L'_{max} - L'_{min})/(L'_{max} + L'_{min})$, it can then be established that the function G represents precisely the number of grey shades.

It follows from the formula that in the determination of the GSFP the higher modulation values will carry more weight than the lower ones. Task himself remarks that this appears to go against all psychophysical indications. In fact, all other similar resolution measures in the literature describe a reverse tendency, in which lower modulations carry more weight than the higher ones.

2.2.4 Integrated contrast sensitivity

Another way of incorporating the contrast modulation threshold of the visual system in a resolution measure is presented by Van Meeteren (22). He considers the contrast sensitivity (the reciprocal value of the contrast modulation threshold) as an MTF which gives a description of the final system component, namely the eye. The total MTF which is thus calculated as the product of the system MTF and the contrast sensitivity is then integrated over all spatial frequencies:

$$ICS = \int_0^{\infty} MTF_s(w) \cdot C_{v_1}(L, \frac{2\pi a}{360} \cdot w)dw,$$

which results in the "integrated contrast sensitivity". In this ω is linked to w via the viewing distance a. The contrast sensitivities $C_{v1}(L,\omega)$ are determined by Van Meeteren for luminances L between 10^{-4} and 10 cd/m^2. According to Van Meeteren the ICS is directly related to the photon flux detected by the eye.

In comparison with the MTFA the ICS reflects more emphatically the dependence on the contrast modulation threshold. Whereas contrast modulation thresholds of 0.1 and 0.01 produce no appreciable difference in the case of the MTFA, this factor of 10 is expressed fully in the ICS.

2.2.5 Power law model

In imitation of the many power law stimulus-response models of Stevens (31), Hunt and Sera (18) also tried to accommodate picture quality in such a model. One of the parameters they chose for picture quality was resolution. The measure for it was for them in fact the stimulus (power law stimulus):

$$PLS = \int_0^\infty \mathcal{F}(\log(PSF(x,y))).\ C_{v_1}(L,\ \frac{2\pi a}{360}\ .\ u,\ \frac{2\pi a}{360}\ .\ v)du dv.$$

The formula is therefore based on the Fourier transform of the **logarithm** of the point spread function of the system. The use of the logarithm is based on Weber's law. The contrast sensitivity $C_{v1}(L,\mu,\nu)$ is taken from Granger and Cupery (15) (see section 2.2.2). The spatial frequencies of the eye μ and ν are converted with the aid of the viewing distance a to those on the image u and v.

A number of photos made with the aid of digital image processing equipment are judged in terms of quality in a magnitude scaling experiment. Although it would appear from the title of their article that Hunt and Sera are searching for a quality measure in a non-task-oriented environment and the scaling experiment ties in with this, it should be borne in mind that the experiment was executed on the basis of aerial photos in a military (detection-oriented and hence performance-oriented) environment. The averaged responses R in the experiment appear in fact to correlate via a power law with the PLS, which is expressed in the following formula:

$$R = k\ .\ (PLS - PLS_0)^{\beta}.$$

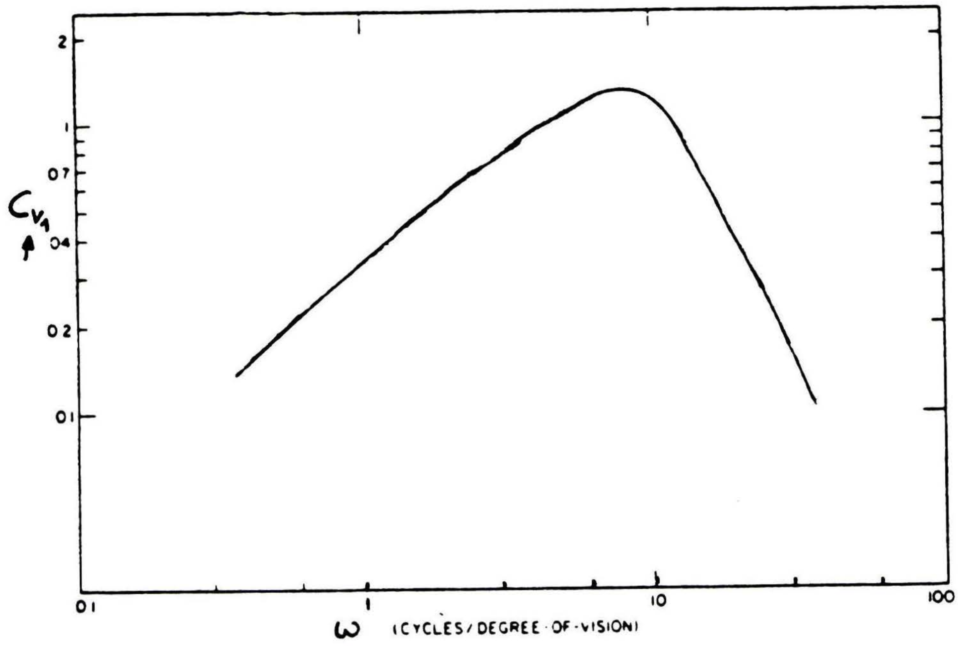

Figure 9: Contrast sensitivity curve
This contrast sensitivity curve was determined by Davidson and is used by Cohen and Gorog in their visual capacity resolution measure. It is not specified at what luminances and under what conditions the curve was measured.

where β is a constant (approximately 0.4) and PLS_0 and k are dependent on the signal-to-noise ratio. In fact, Hunt and Sera are describing a system here in which a compressing non-linearity occurs twice: the logarithm in the calculation of PLS and the exponent β which is smaller than 1.

2.2.6 Visual capacity

Following Shannon (29), who gives a measure for the information capacity of an electrical communication channel, Cohen and Gorog (10) developed the visual capacity. By analogy the visual capacity can be considered as the total number of edges that can be perceived at a given distance a from the image of width d:

$$VC = \frac{360}{2\pi}\,\frac{d}{a} \cdot \frac{1}{\theta_e(a, MTF_s)}$$

The factor d/a is the opening angle which subtends the image for the observer and is proportionate to its size on the retina. The term $360/2\pi$ converts this opening angle from radials to degrees. θ_e(also expressed in degrees) can be considered as the angle which the imaging system and eye together need for the rendering of a step function. The visual capacity thus presents the maximum perceivable number of edges or contours for a certain imaging system and a certain viewing distance. Cohen and Gorog indicate a division by two in this formula: the term $360 \cdot d/2\pi a$ gives in fact the total quantity of information of the image; the angle θ_e is considered as a measure for resolution (sharpness) and is essentially dependent on viewing distance and system MTF:

$$\frac{1}{\theta_e(a, MTF_s)} = 2\,K \int_0^{\infty} \left|\, MTF_s\left(\frac{360}{2\pi a}\cdot\mu\right) \cdot C_{v_1}(L,\mu)\, \right|^2 d\mu.$$

The integration only takes place in the horizontal dimension u, because it is intended to apply the VC specifically to raster-scanned television screens. The function $C_{v1}(L,\mu)$ is the contrast sensitivity curve for which data of Davidson (12) (see figure 9) are used. It is not indicated at what luminances and under what other conditions this contrast sensitivity curve applies. This curve is normalised by the dimensionless constant K so that the VC supplies precisely the number of TV lines at the optimal viewing distance. One of the qualities of the VC is that this optimal viewing distance also genuinely exists and can be found by maximising VC in relation to a.

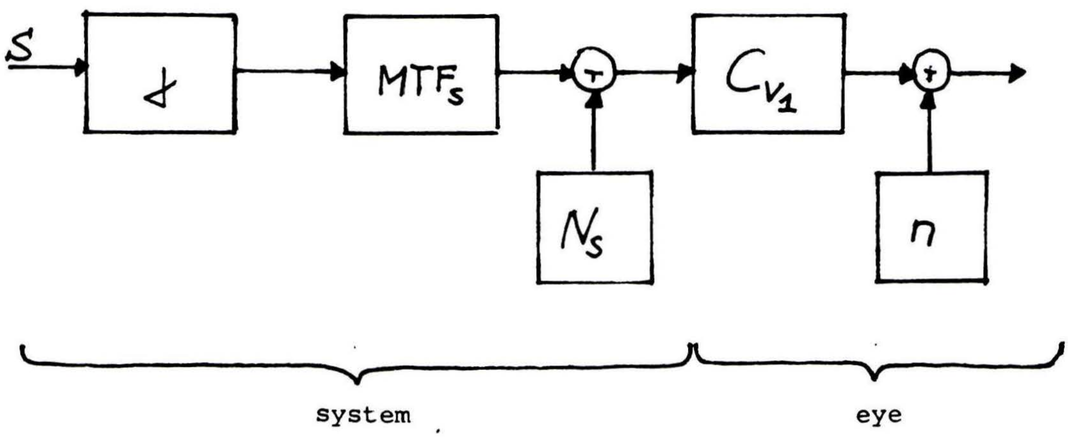

Figure 10: Diagram for calculation of SNC

The calculation of the resolution measure θ_e calls to mind the noise equivalent passband of Schade (section 2.1.3), with the contrast sensitivity curve as the MTF of an extra system component.

2.2.7 Signal-to-noise ratio criterion

Nelson (23) also tries to apply the information theory of Shannon (28) to an imaging system, notably to photography. In contrast to Cohen and Gorog (section 2.2.6) however, he takes account explicitly of the noise in the imaging system in the form of the Wiener noise spectrum $N_s(w)$. The visual system of the observer is in the first instance left out of consideration. His "information transmission capacity" then becomes:

$$ITC = \pi \int_0^\infty w \, log_2 \left(1 + \frac{|S(w)|^2 \gamma^2 MTF_s^2(w)}{N_s(w)}\right) dw.$$

where γ is the exponent in the relationship between the light intensities before and after the imaging system: $L=L'^{\gamma}$. In the formula an attempt is made to approximate this by a multiplication by γ. The numerator of the fraction in the formula is in fact the signal for the eye, calculated from the product of the spectrum content S(w) and system MTF; the denominator comprises a noise term which relates to the imaging system. To this ratio a value of 1 is added so that a logarithm can be taken in all cases.

Sometime later Nelson modified his ITC in such a way that the influence of the visual system was also taken into account. The model which he uses is sketched in figure 10. From this there emerges a new measure, the "signal-to-noise criterion" and is described by, among others, Higgins (17):

$$SNC = 100\sqrt{a/a_r},$$

where

$$a = \int_0^\infty log \left(1 + \frac{|S(w)|^2 \gamma^2 MTF_s^2(w) C_{v_1}^2(\frac{2\pi a}{360} \cdot w)}{b(\frac{2\pi a}{360} \cdot w) \cdot (N_s(w) C_{v_1}^2(\frac{2\pi a}{360} \cdot w) + n)}\right) dw,$$

and the reference value a_r is thereby calculated for the hypothetical case that $MTF_s(w) = 1$ for all w. Why the term w has now disappeared from the integrand of the ITC is not stated; it probably has to do with transformations due to the fact that the ITC has a two-dimensional nature, whereas the SNC only attempts to describe the resolution in one dimension.

In addition to the visual contrast sensitivity $C_{v1}(\omega)$, several other visual quantities also play a role. The term n has to do with the biological noise in the eye but is not indicated further. The critical noise bandwidth $b(\omega)$ has approximately the size of an octave around the frequency ω, according to measurements by Stromeyer (32). In the SNC formula the eye is considered as the final system component. All output of the imaging system is multiplied by the contrast sensitivity curve, which is taken the MTF of the visual system. However, the eye also produces noise itself, which is added to the noise of the imaging system.

2.2.8 Just noticeable differences (JND model)

The JND (just noticeable differences) model of Carlson and Cohen (6) differs slightly from the resolution measures already described due to the fact that it is on the one hand based on an extended model of the visual system and on the other hand because a real measure of resolution can only be deduced in the second instance from the results of the model. The basis for the JND model is the quadratic detection of differences in contrast modulation. The distinction between a modulation m_0 and a slightly larger modulation m is still just detected if the following formula is satisfied:

$$m^2 - m_0^2 = km_0^2 + m_t^2$$

The difference between the two modulations $\Delta m_0 = m - m_0$ is now called 1 JND. The size of the JND is dependent on the output modulation m_0 and the threshold modulation m_t, which is measured if $m_0 = 0$, and in fact is a measure for the visual noise in the eye. The factor k ensures that for larger values of m_0 the model satisfies Weber's law: $\Delta m_0 / m_0 = k/2$.

Carlson and Cohen assume the presence of independent frequency-specific channels in the visual system, on the basis of measurements by Sacks (24). Their model describes the visual system with seven channels with channel frequencies of between 0.5 and 48 periods/degree, spaced at logarithmically equal distances. The channels have a width $\Delta\omega$ equal to approximately 1 octave around the channel frequency ω. Within each channel ω the quadratic detection model now applies, which results in:

$$|S(w)|^2 \frac{2\pi\Delta\omega}{a}\left(MTF_{s,b}^2(w) - MTF_{s,a}^2(w)\right) =$$

$$k\left(N_s(w)\frac{2\pi\Delta\omega}{a} + N_v(\omega)\Delta\omega + |S(w)|^2 \frac{2\pi\Delta\omega}{a} MTF_{s,a}(w)\right).$$

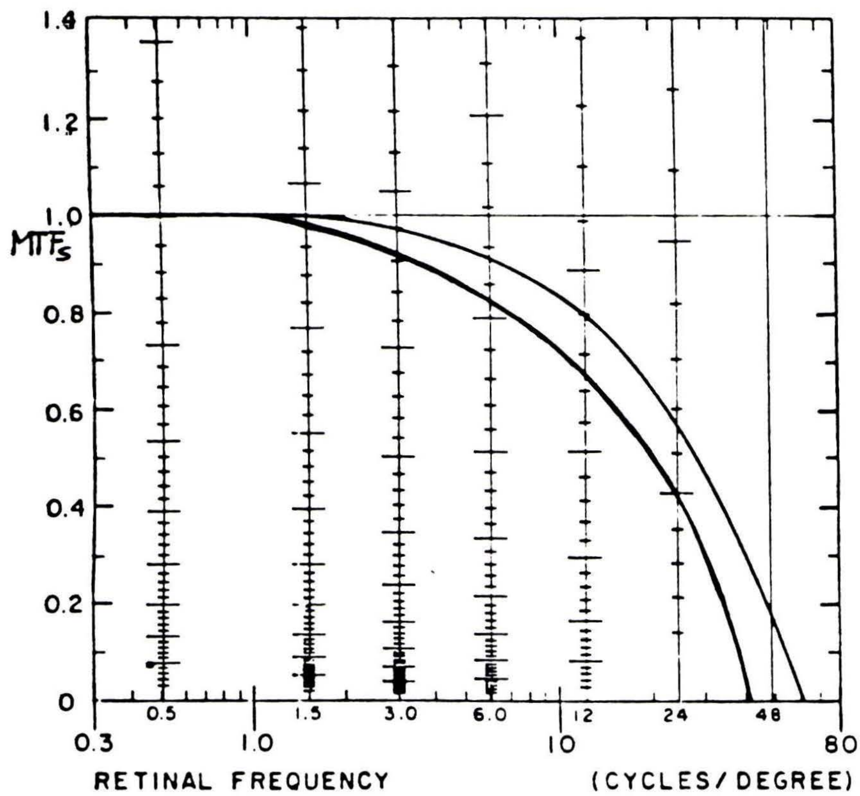

Figure 11: Discriminable Difference Diagram

Two possible MTFs are drawn in. The difference between them in JNDs can be read off directly in each frequency channel.

Weighted over the spectrum of the scene content S(w) a comparison is made between two possible MTFs of the system, namely $MTF_{s,a}$ and $MTF_{s,b}$. In channel ω they differ by precisely 1 JND when the above formula is satisfied. The noise which influences the detection is in this formula formed by both the noise of the image $N_S(w)$ and by the noise per unit of retinal frequency of the eye $N_V(\omega)$. The spatial frequency on the retina ω and that on the image w can again be
transformed into one another at a known viewing distance a: $w=\frac{360}{2\pi a}\cdot\omega$.

As the scene content is included in the formula, the result of the calculation is thus picture-dependent. This is less desirable, and an approximation has been sought which would be satisfactory for a large proportion of the possible scenes. Carson (7) finds for this:

$$|S(w)|^2 = \frac{(\Delta L)^2}{\pi} \frac{\sin^4(wd/4)}{2\pi w^2},$$

where ΔL is the luminance difference in the scene and d the width of the image.
With the aid of this estimated scene content and with the required parameters of the imaging system as the input, the so-called discriminable difference diagrams can be calculated. These diagrams, see for example figure 11, are best considered as high-quality graph paper. When two different possible MTFs are drawn in for the imaging system, the number of JNDs difference can be read off directly in each channel.

There are various ways in which the different frequency channels can work together. It is possible that with a changing MTF a difference is detected on the criterion that the difference value of 1 JND is obtained in one of the frequency channels. Another point at which the detection could occur is when the (fractional) JNDs of all channels added up together give the value of 1 JND. An experiment has shown however that on this first assumption just an upper limit is found for the measured values and on the second assumption just a lower limit. A third possibility, the summation of detection probabilities per channel, produces an intermediate, optimal prediction of the measured values. A possible theoretical reason for the two methods of summation is on account of statistical considerations, but this is nowhere mentioned explicitly.

A measure of resolution for a system MTF could be derived from the JND model by comparing it with the ideal MTF which is 1 for all frequencies and by simply adding up the number of JNDs difference in all channels. Although Carson and Cohen state that they have their doubts as to the correctness of this method, they do use it for the evaluation of picture quality (8).

2.2.9 Square root integral

One of the positive characteristics of the JND model from the previous paragraph is that it has a sound psychophysical basis; the disadvantage is however that the method of calculation is too complicated for it to be a handy measure of resolution. With his square root integral Barten (2) attempts to find a compromise. He uses the Carlson and Cohen JND model (paragraph 2.2.8) as a basis. Specifically, he uses the estimation of the difference modulation (Δm) required for detection at a high modulation value m, where an asymptotic relationship applies: $\Delta m=km$. On the basis of measurements by Carlson and Cohen, Barten modifies this formula to $\Delta m=km^n$, and also extrapolates this equation to the lower modulation values. In that case it can easily be calculated how many JND units separate the modulation m from the modulation 0:

$$j = (m/m_t)^{(1-n)},$$

where j denotes the number of JND units and m_t the threshold modulation. Barten now makes a further assumption, one which Carlson and Cohen did not wish to make, namely that all JNDs in all frequency channels can be added up to give a measure for the 'visual resolution'. This results in an integration over the frequency domain, and on logarithmic basis, because in the Carlson-Cohen model too the frequency channels lie at logarithmic distances from each other.

$$j_{sum} = K \int_0^{w_{max}} \left(\frac{MTF_s(w)}{m_{d,v_1}(L, \frac{2\pi a}{360} \cdot w)} \right)^{(1-n)} d(log(w)).$$

In this integral the absolute modulation m is replaced by the MTF of the imaging system $MTF_s(w)$ which is a relative measure. It is not to be expected therefore that the integral will actually produce a sum of JNDs, in view of the fact that the contrast modulation m_0 of the target would also have to be included in the formula. The expression can however be considered as a resolution measure of the MTFA type.

For the contrast modulation threshold $m_{d,v1}(L,\omega)$ Barten uses an extrapolation of data of Van Meeteren (22). The spatial frequency on the retina ω is of course again converted to that on the image w with the aid of the viewing distance a. It is unclear whether the integration limit w_{max} is determined by the equation $MTF_s(w_{max})=m_{d,v1}(L,\frac{2\pi a}{360}.w_{max})$ or by the highest spatial frequency which is present in the scene content.

In any case the constants K and 1 - n of this resolution integral are found by application of the expression to some of the measurement data of Carlson and Cohen. For 1 - n an optimal value of 0.5 was found and for K the value of 1/ln(2) appeared to be satisfactory. This latter value is in conformity with the expectations insofar as it represents precisely the reciprocal of the width of frequency channel. In this way when $MTF_s(w) = m_{d,v1}(L\frac{2\pi a}{360}.w)$ in a single frequency channel of the width ln(2) (and moreover when $m_0 = 1$) exactly the value expected of 1 JND recurs for the integral. The formula for this square root integral then becomes:

$$SQRI = \frac{1}{ln(2)} \int_0^{w_{max}} \sqrt{\frac{MTF_s(w)}{m_{d,v_1}(L, \frac{2\pi a}{360} \cdot w)}}\, d(log(w)).$$

The division by the contrast modulation threshold $m_{d,v1}(L, \omega)$ in fact comes down to multiplication by the contrast sensitivity $C_{v1}(L,\omega)$. The ultimate formula of the SQRI then corresponds in terms of structure with, for example, the SQF of Granger and Cupery (section 2.2.2) or the $MTFA_{log-log}$ of Synder (section 2.2.3).

The SQRI appeared to fit poorly to the results of a small number of discrimination measurements by Barten.

3 COMPARATIVE STUDIES

Many of the resolution measures mentioned in the previous chapter were presented without there being any real validation of the measure in question. Some authors accompanied their proposals with a number of perception experiments, but these were either limited in scope, poorly described or in some cases completely failed to support the proposed resolution measure. A few attempts were made to compare the various measures of resolution. Sometimes such a comparison coincided with the introduction of a new resolution measure. Four such articles are dealt with in this section.

A number of qualifying remarks needs to be made in this context. First of all, the author often has a specific purpose in mind for the resolution measure. This might be, for example, the description of imaging equipment for task-oriented objectives such as military detection or for example amusement purposes or to describe the picture material itself. Secondly, it often happens that the author slightly modifies the resolution measures from the literature to suit his own views. This sometimes has to do with the aim the author has in mind for the resolution measure. A third observation concerns the experimental configuration. This also varies, i.e. between detection tasks and quality judgments, but unfortunately it was not often considered in what way the results of the two methods are linked. All in all, it is therefore to be concluded that the comparative experiments to be described are not necessarily comparable themselves.

3.1 Experiments by Snyder

Snyder (30) describes three experiments which all have the aim of validating the MTFA as a measure of resolution, but in which a number of other measures are also involved.

EXPERIMENT 1:

- Material: Nine aerial photographs of military objects. These were copied per scene under various conditions, causing 32 different stimuli to arise, varying in MTF_s, contrast modulation and grain structure (static noise).

Table 1: Correlations of various resolution measures with the values on the subjective quality scale.

The correlations are ordered per scene. The average correlation per scene is calculated via the mean of the respective z scores.

Physical variable	Scene number									Mean* r
	1	2	3	4	5	6	7	8	9	
MTFA (linear)	0.921	0.927	0.900	0.925	0.935	0.919	0.919	0.920	0.913	0.920[†]
Modulation	0.220	0.641	0.511	0.618	0.680	0.699	0.497	0.698	0.632	0.576
MTF	0.698	0.529	0.580	0.660	0.579	0.608	0.697	0.469	0.542	0.601
Granularity	−0.543	−0.632	−0.618	−0.450	−0.516	−0.428	−0.505	−0.589	−0.577	−0.543
MTFA (log-log, 2 cycle)	0.666	0.863	0.866	0.821	0.874	0.890	0.749	0.902	0.876	0.846
MTFA (log-log, 2 cycle)	0.768	0.923	0.923	0.867	0.920	0.921	0.824	0.941	0.920	0.900
Acutance (SMT)	0.599	0.448	0.526	0.568	0.564	0.599	0.625	0.440	0.602	0.555

- Subjects: 36 experienced 'photo interpreters': these are people whose job it is to derive military information from aerial photographs.

- Method: The subjects were each presented with 256 pairs. Each pair comprised two versions of the same scene. The subject had to indicate one of the two as the best quality with a view to information extraction.

- Processing: From the percentage preference for various stimuli a subjective quality scale was constructed. The position of the various stimuli on this scale was correlated with a number of resolution measures, on the assumption of a linear relationship (Pearson's product correlation; for the description of this and other statistical parameters see for example Guildford (16)).

- Results: A summing up of the correlations established with a number of measures of resolution is found in table 1. The correlations are calculated per scene. The average value for all scenes is calculated by averaging Fisher's z scores of the 9 scene correlations and from this calculating the average correlation coefficient r.

It emerges from the table that the three measures based on MTFA are highly satisfactory. The difference between the two $MTFA_{log-log}$ measures is not known. The fact that the $MTFA_{log-log}$ also attains such a high correlation is important because this measure is to some extent related to measures such as ICS, PLS, VC and SQRI. These are all measures in which the system information is divided by a visual threshold curve.

In terms of performance the (SMT) acutance falls far short of the MTFA measures. Other measures such as Modulation, MTF and Granularity only achieve very low correlations. This is to be anticipated because they all describe only one of the three dimensions in which the stimuli are varied. From the correlations it can be further deduced that all three dimensions contribute to a more or less equal extent to the impression of quality.

Table 2: Correlations between MTFA, percentage errors and categorical scaling.
These are ordered per scene. $\bar{r}$ is determined by averaging the z scores per scene; r_m is the correlation determined by taking all measurements together.

	Scene									$\bar{r}$	r_m
	1	2	3	4	5	6	7	8	9		
Performance/MTFA	0.69	0.66	0.80	0.65	0.78	0.55	0.84	0.86	0.46	0.72	0.93
Performance/rank	0.71	0.67	0.89	0.60	0.80	0.42	0.78	0.76	0.42	0.70	0.96
MTFA/rank	0.90	0.87	0.90	0.93	0.94	0.87	0.92	0.86	0.83	0.90	0.97

EXPERIMENT 2:

. Material: The same material was used as in the experiment described above.

. Subjects: 384 experienced photo interpreters.

. Method: Each subject was presented with one of the 288 possible stimuli. He had to do two things with it:

1. Position the stimulus on a nine-point categorical scale. The criterion to be used was the picture quality for information extraction purposes.

2. Answer a number of multiple choice questions relating to the scene content.

. Processing: For the categorical scaling the average of all subjects was calculated for each stimulus. From the multiple-choice questions the number of errors was recorded for the total of all subjects. Correlations between these two quality-describing parameters and with the MTFA were determined on the assumption of a linear relationship (Pearson's product correlation).

. Results: Table 2 gives the correlations between MTFA, percentage errors (performance) and categorical scaling (rank). The averaging over the 9 different scenes was done in two ways. r denotes the correlation, associated to the z score which arises by averaging the z scores of the 9 correlations per scene. r_m denotes the correlation which arises when no distinction is made between the scenes.
The table reveals first and foremost a very high correlation between all three quality-describing parameters for the total of all scenes. Per scene however, a significantly higher value is found for the correlation between MTFA and categorical scaling than for the other two correlations. As a quality-describing parameter the percentage errors also appears to be much more sensitive to the scene content.

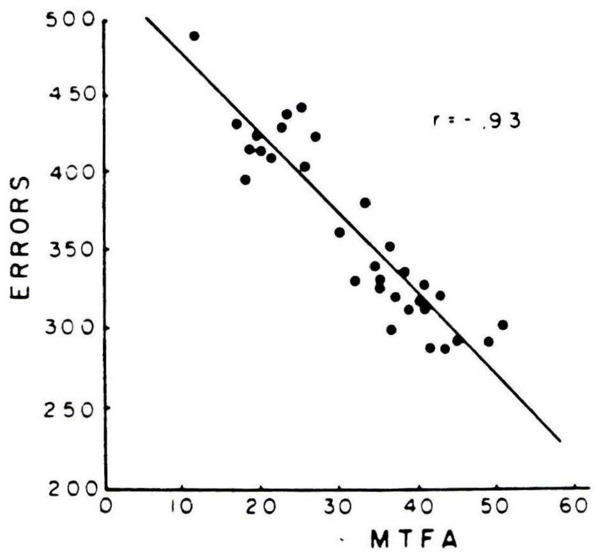

Figure 12: Scatter diagram of numbers of errors against MTFA.

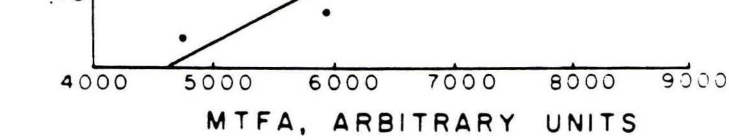

Figure 13: Relationship between percentage correct scores and MTFA

This leads to the somewhat strange conclusion that both the subjects in their quality scaling and the MTFA resolution measure show the same deviations for the same stimuli when they try to predict the performances of the subjects. In other words: the MTFA correlates better with the prediction of the subject as regards performance than with the performance itself. A possible reason for this might be that the parameter percentage errors is not linear but perhaps correlates in another (for example quadratic) manner with the two other parameters, in which case the Pearson's product would appear to underestimate the measure of correlation. This cannot however fully explain the scene dependence. Nor do a scatter diagram for the MTFA and the number of errors (figure 12) point clearly in this direction.

EXPERIMENT 3:

- Material: Aerial photographs with 25 unknown targets were shown on a 945-line, 16 MHz monitor with a fixed MTF. The signal-to-noise ratio of the video signal for the monitor was set at five different values.

- Subjects: For each signal-to-noise ratio there was a group of 11 subjects, probably again photo interpreters.

- Method: At a viewing distance of 40 cm the subject was asked to recognise the 25 targets in a prescribed sequence. It was also at this distance that the 5 $m_{d,v+s}(L,w)$ curves were measured for each signal-to-noise ratio.

- Processing: The percentages of correct recognitions were calculated and correlated linearly with the MTFA values.

- Results: The results are plotted in figure 13. The correlation found is very reasonable, although a quadratic fit would probably lead to even better results. It should be stressed again that the variation in MTFA values arose through the use of different detection curves $m_{d,v+s}(L,w)$ as a consequence of the different signal-to-noise ratios of the video signal, and not through variation of the system MTF.

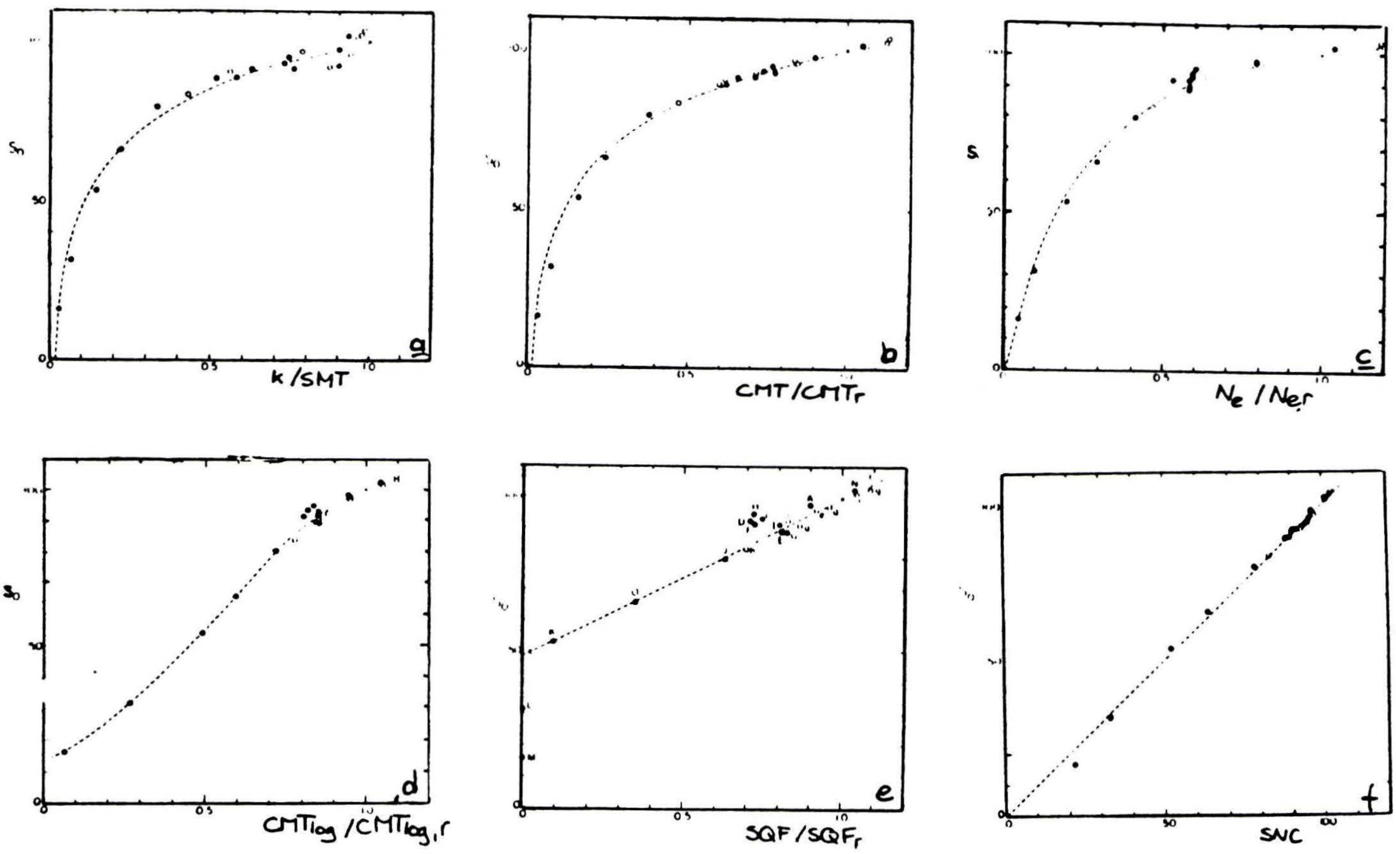

Figure 15: Correlations of sharpness with different resolution measures.

The vertical axis is the sharpness scale; on the horizontal axes the various (normalised) resolution measures are plotted. The subscripts r indicate normalisation values which are found for the resolution measure when MTF_s is 1 for all frequences.

a) resolution measure k/SMT. The constant k relates to the sharpness limit of the visual system.

b) resolution measure CMT/CMT_r. Higgins only took the result of the integral for the calculation of the CMT and left out all other operations.

c) resolution measure $N_e/N_{e,r}$. It is however probable that Higgins has included the contrast sensitivity of the eye in this measure also as an MTF of the system.

d) resolution measure $CMT_{log}/CMT_{log,r}$. This measure is the same as the CMT, only now there is an integration over the frequencies on logarithmic basis.

e) resolution measure SQF/SQF_r

f) resolution measure SNC

3.2 Experiments by Higgins

Higgins (17) compared a large number of measures of resolution in an otherwise summarily described experiment.

- Material: Photos of four different scenes were used, probably normal complex scenes, size 10 x 10 cm^2. These were printed with a broad scale of 22 different MTFs with normal, but also sometimes widely differing forms (see figure 14).

- Subjects: 20 people, whose background was not specified, took part in the experiment.

- Method: In the description of the experiment there is only mention of a "subjective evaluation", but it is not clear what test procedure this is based on, nor what criterion is used. In this latter respect Higgins seems to consider quality and sharpness as completely synonymous.

- Processing: After "using appropriate psychological scaling procedures", the subjective data finally produce a 100-point sharpness scale.

- Results: Higgins finally plotted the sharpness scales as a function of a large number of resolution measures. In doing so he normalised, insofar as necessary, all existing measures of resolution. He took as the normalisation factor the value of the resolution measure found when the MTF_s for all frequencies is 1. The six respective graphs are shown in figure 15. It can be deduced from this that the CMT acutance and the SNC are the most suitable as measure of resolution. Both measures describe the visual system as an additional component of the imaging system. Higgins states in the accompanying text that the deviations in the case of the other measures of resolutions are essentially attributable to the more bizarre MTF shapes.
  A further experiment, described if possible even more summarily, was carried out to decide between CMT and SNC. For this purpose the photographs were provided with three noise levels. This was done for

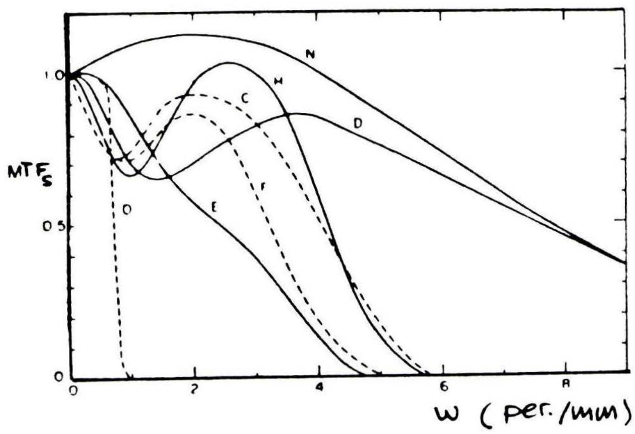

Figure 14: Seven of the 22 MTFs used
These have the most bizarre shapes: the remaining 15 curves were more Gaussian.

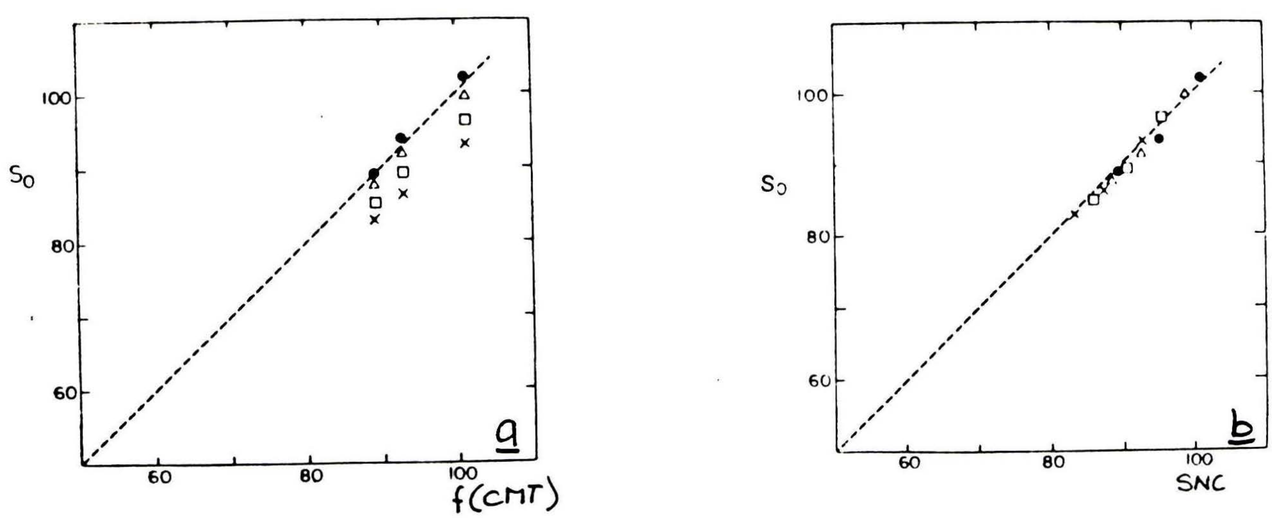

Figure 16: Correlations of two resolution measures with sharpness in the presence of noise.
The vertical axis is again the sharpness scale; on the horizontal axis are two resolution measures.
a) resolution measure CMT.
b) resolution measure SNC.

three fixed MTFs. The results for these 12 stimuli ((3 + 1) x3) were processed in the same way as for the above experiment and again put into graph form, see figure 16. The SNC measure appears to be better for describing the influence of the noise, which is not surprising, since in the formula of the SNC a noise term is explicitly included.

3.3 Experiments by Task

Task (33) subjects the MTFA to a number of modifications in a comparative study. The MTFA is here used in CRT display research and it is probably assumed that the signal-to-noise ratio is large enough to replace the threshold curve for imaging and visual system $m_{d,v+s}(w)$ by the contrast threshold of the visual system alone $m_{d,v1}(L,\omega)$ or $m_{d,v2}(L,\omega)$. In fact, it is $MTFA_{lit}$ which is referred to here.

- Material: Six toy models of tanks and other implements of war were placed against a black-and-white checked background, and photographed from four angles. The purpose of the checked background was to reduce the silhouette information. A 20-second film was made from the photos by zooming in with a film camera. The film had to simulate the view from an approaching vehicle. The films were stored in an image-processing system and shown on a monitor. The video bandwidth was 6.0, 1.0 or 0.4 MHz, so that only the horizontal resolution was varied. The maximum contrast could be set to a value of 50, 10 or 3. The maximum screen luminance was always 230 cd/m^2. There were therefore 9 display conditions, of which the MTF was always accurately determined.

- Subjects: 36 women and 36 men, all with a vision of 20/20, corrected or uncorrected.

- Method: Prior to the session detailed photos of all six war vehicles were studied by the subjects. During the session the subject had to stop the zooming in of the film by pressing a button at the point when he was sure enough to be able to say which vehicle it was. The viewing distance was 70 cm. The subject received feedback from the

Table 3: Average opening angle of the target at the moment of recognition for the 9 display conditions.

CONTRAST RATIO	BANDWIDTH (MHz)		
	0.4	1	6
50:15	3.9 deg	2.2 deg	1.9 deg
50:5	3.6	1.9	1.6
50:1	3.4	2.4	1.5

Table 4: Correlations of different resolution measures with the average opening angle.

	FOM	Correlation With Performance
1	Log BLMTFA	-0.948
2	Log LLMTFA	-0.932
3	Log Suprathreshold Resolution	-0.911
4	Log MTFA	-0.905
5	Suprathreshold Resolution	-0.888
6	MTFA	-0.811
7	Log Limiting Resolution	-0.786
8	GFP	0.781
9	Limiting Resolution	-0.764
10	GFP-Log	-0.750

experimenter as to the correctness of his response. In the case of a correct recognition the opening angle at the stopping point was taken as a measure for the display quality. Every subject had 48 viewings under a single display condition.

. Processing: For each one of the nine display conditions the mean was calculated of all opening angles measured in the case of correct recognition. These means are shown in table 3. No further comments were made regarding the slightly strange value of 2.4 degrees at a video bandwidth of 1 MHz and a contrast value of 50.

. Results: For a number of resolution measures the correlations were calculated with the average opening angles found. This was probably done on the assumption of a linear relationship. The correlations were also calculated for the logarithms of the resolution measures. This was done because it cannot be automatically assumed that the average opening angle correlates linearly with a possible measure of resolution, particularly since it is not known what sort of influence the zoom speed has on this. The various correlations are shown in table 4. The resolution measure "limiting resolution" is that spatial frequency at which the curves of MTF_s and $m_{d,v1}$ cross each other (thus in fact the upper integration limit of the MTFA). The resolution measure 'suprathreshold resolution' is similarly that spatial frequency w at which the values $MTF_s(w)$ and $m_{d,v2}(\frac{2\pi a}{360}.w)$ are the same.

Neither of the two measures which turned out to be the best contain information on the behaviour of the MTF at frequencies smaller than 2 periods/degree (the lower limit of integration). The fact that this information does not appear to play any role in this experiment is to be explained by the zooming-in movement of the picture cycle. The low-frequency information has already been clearly visible for some time and very probably long since processed by the subject. At the points 1 and 3 are measures of resolution which are based on $m_{d,v2}(\omega)$, the threshold curve for discrimination between a sine and a block raster. This finding ties in with the discrimination task of the experiment. As the fourth, the logarithm of the $MTFA_{lit}$ also seems to correlate well with the

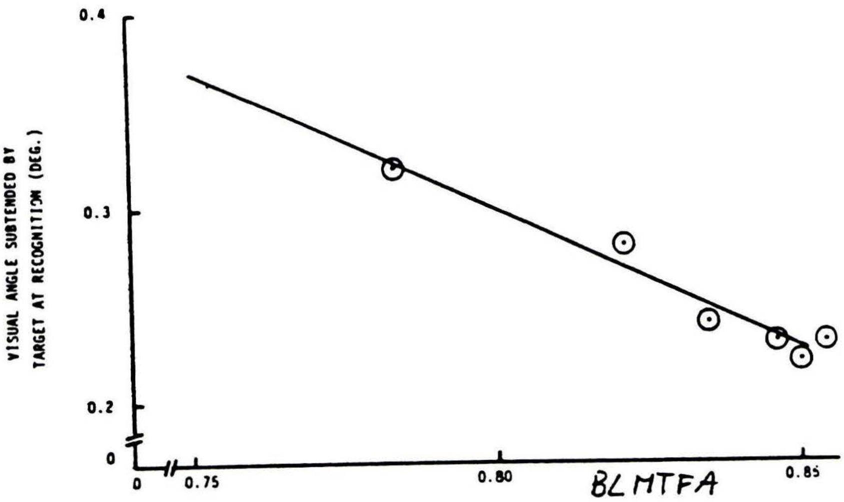

Figure 17: Correlation of BLMTFA with average opening angle.

Table 5: Resolution measures used by Beaton.
The following are shown: the name used by Beaton, Beaton's calculation method and, where possible, the name which is used in this report. Beaton's notation of the physical parameters is indicated under the table.

Metric	Expression	Report name
PEP	$\Delta\omega\Delta\nu\Sigma\Sigma\ S^2(\omega,\nu)$	N_e
EW	$S(0,0)/\Delta\omega\Delta\nu\Sigma\Sigma\ S(\omega,\nu)$	W_e
SSF	$\Delta\omega\Delta\nu\Sigma\Sigma\ S(\omega,\nu)(\omega^2+\nu^2)$	
MTFA	$\Delta\omega\Delta\nu\Sigma\Sigma\ S(\omega,\nu)-T(\omega,\nu)$	$MTFA_{lit}$
GSFP	$\Delta\omega\Delta\nu\Sigma\Sigma\ G\{S(\omega,\nu)-T(\omega,\nu)\}$	GSFP
where G{•} denotes gray shade transform		
ICS	$\Delta\omega\Delta\nu\Sigma\Sigma\ S(\omega,\nu)$	ICS
Q3	$\dfrac{\Delta\omega\Delta\nu\Sigma\Sigma\ S^2(\omega,\nu)}{1+\{\Delta\omega\Delta\nu k\Sigma\Sigma\ W(\omega,\nu)M_e^2(\omega,\nu)\}}$	
PMR	$\Delta\omega\Delta\nu\Sigma\Sigma\ M_s(\omega,\nu)F(\omega,\nu)/M_e(\omega,\nu)$	
MSE	$\dfrac{\Delta\omega\Delta\nu\Sigma\Sigma\ \{F_0(\omega,\nu)-F(\omega,\nu)\}^2}{\Delta\omega\Delta\nu\Sigma\Sigma\ F_0(\omega,\nu)}$	
PMSE	$\dfrac{\Delta\omega\Delta\nu\Sigma\Sigma\ \{S_0(\omega,\nu)-S(\omega,\nu)\}^2}{\Delta\omega\Delta\nu\Sigma\Sigma\ S_0(\omega,\nu)}$	
IF	$1-\dfrac{\Delta\omega\Delta\nu\Sigma\Sigma\ \{S_0(\omega,\nu)-S(\omega,\nu)\}^2}{\Delta\omega\Delta\nu\Sigma\Sigma\ S_0^2(\omega,\nu)}$	IF
SC	$\Delta\omega\Delta\nu\Sigma\Sigma\ S^2(\omega,\nu)/\Delta\omega\Delta\nu\Sigma\Sigma\ S_0^2(\omega,\nu)$	SC
CQ	$\dfrac{\Delta\omega\Delta\nu\Sigma\Sigma\ S_0(\omega,\nu)S(\omega,\nu)}{\Delta\omega\Delta\nu\Sigma\Sigma\ S_0^2(\omega,\nu)}$	CQ
IC	$\Delta\omega\Delta\nu\Sigma\Sigma\ \log_2\{1+S(\omega,\nu)/S_0(\omega,\nu)\}$	

where $F(\omega,\nu)=|I(x,y)\exp\{-j2\pi(\omega x+\nu y)/N\}|$ denotes the 2-dimensional discrete Fourier transform of image $I(x,y)$, $S(\omega,\nu)=M_s(\omega,\nu)M_e(\omega,\nu)F(\omega,\nu)$ denotes the displayed modulation spectrum, $S_0(\omega,\nu)$ refers to a non-degraded image, ω and ν refer to spatial frequencies in units of cycles per pixel, $M_s(\omega,\nu)$ denotes the 2-dimensional MTF of the imaging system, $M_e(\omega,\nu)$ denotes the normalized 2-dimensional MTF of the human visual system, $T_e(\omega,\nu)$ denotes the 2-dimensional contrast threshold function of the human visual system, and $W(\omega,\nu)$ denotes the 2-dimensional Wiener noise spectrum.

average opening angle. The GSFP measure is significantly less satisfactory, which Task attributes the fact that the grey shade transformation implemented here emphasises precisely the higher modulation values instead of the values which lie closer to the threshold modulation.
Because the BLMTFA measure was constructed on the basis of the aforementioned experiment, a second experiment was carried out in order to confirm the validity of the BLMTFA. Six different display conditions were constructed by varying the viewing distance at a constant (maximum) video bandwidth and contrast. The subjects in this case were experienced, but of unknown number, and they conducted the same test as in the aforementioned experiment. In this case too, a very high correlation (r=-0.977) was found between the BLMTFA and the average opening angle on recognition, as can be seen from figure 17.

3.4 Experiments by Beaton

Beaton (3) undertook the analysis of an extensive set of data. This database consists of two halves: one describes the physical data of a number of scenes; the other contains performance and scaling data of a number of photo interpreters. How the data were obtained is explained in detail below. Beaton intended to use these data in order to find a measure of quality, not so much for the imaging equipment, but for the image itself. For this reason he incorporated explicit information on the scene content, in particular the spectrum of the scene S(u,v), in all measures of resolution where this had not yet been done. All measures were for this purpose also transformed into a two-dimensional description. The visual system was included in all measures of resolution as an extra component of the imaging system. It is not known however what data Beaton used for the contrast sensitivity. An explanation of the measures of resolution used by Beaton, the manner in which he calculated them and their origin is given in table 5.

- Material: Ten aerial photographs of military bases were entered into an image-processing system, with a resolution of 4096x4696 pixels and 6 bits grey shades. These were then degraded in two ways. The pictures underwent low-frequency filtering with

Table 6: Correlations of resolutions with quality judgments.
For type I the correlation was calculated for all data together; type II represents the mean of all correlations per scene.

Metric	Type I (N = 250)	Type II (N = 25)
Hard-Copy Display Condition		
PEP	0.241	0.272
SSF	0.354	0.528
EW	-0.273	-0.407
MTFA	0.315	0.840
GSFP	0.461	0.626
ICS	0.382	0.555
Q3	0.239	0.269
PMR	0.382	0.557
MSE	-0.563	-0.733
PMSE	-0.428	-0.528
IF	0.622	0.868
CQ	0.020	-0.120
SC	0.397	0.548
IC	0.415	0.572
Soft-Copy Display Condition		
PEP	0.445	0.508
SSF	0.463	0.688
EW	-0.490	-0.734
MTFA	0.458	0.921
GSFP	0.618	0.794
ICS	0.575	0.759
Q3	0.440	0.502
PMR	0.575	0.759
MSE	-0.311	-0.368
PMSE	-0.197	-0.199
IF	0.466	0.598
CQ	0.305	0.278
SC	-0.180	-0.186
IC	0.578	0.782

the aid of Gaussian point spread functions with five different widths. Synthetic noise was also added at five different levels (on the basis of the noise which is present on a uniformly exposed film). The thirty resulting pictures were stored with a resolution of 8 bits and reproduced in two different ways. The first method is called soft copy and consists of showing the pictures on a CRT monitor. For the hard-copy facility the pictures were transferred onto film transparencies of 10x13 cm^2.

- Subjects: The subjects were again photo interpreters: 14 in the hard-copy experiment and 15 in the soft-copy experiment.

- Method: The subjects had to fit all 250 pictures into a 100-point category scale. The possibility of extracting information was taken as the criterion for the quality.

- Processing: The values of the measures of resolution were calculated for the various stimuli and subsequently transformed into a scale with a mean of 0 and a spread of 1. With these transformed values the linear correlation was calculated for each measure of resolution with the average scale values.

- Results: The correlations of the various measures of resolution with the scaling values are shown in table 6. The left-hand column in the table (type I) gives the correlations over all stimuli together; the values of the right-hand column (type II) were calculated after the averaging of all correlations per scene. A comparison of the two columns shows immediately that the type II correlations lie significantly higher than the type I correlations. It must be concluded from this that the scaling data show a distinction per scene, at least when they are plotted on the basis of one of these measures of resolution. This may have to do with the fact that Beaton explicitly included the scene information S(u,v) in all measures, because where the resolution is scene-dependent, there is the possibility that precisely this causes the correlation to become scene-dependent.

With a few reservations it can be stated here that the quality judgment of the subjects is not so heavily scene-dependent as the resolution measure itself.

The type I correlations are in general disappointingly low and the ranking of the measures of resolution is so different in the two display conditions that Beaton otherwise leaves the type I correlations out of consideration. In the type II correlations the MTFA is on average the more satisfactory of the two display conditions. Note that not only the scene information is included in the MTFA which Beaton calculates, but also that the $MTF_s(w)$ is multiplied by the contrast sensitivity of the eye. In this way the visual system is taken into account in two ways. The SSF measure, the only one which weighs the higher frequencies more heavily than the lower, only takes a place in the middle the classification. The Q_3 measure, which is the only one which explicitly takes the noise into account, even emerges as the poorest of all.

4 Conclusions and discussion

4.1 Critical remarks

In view of the previous sections and before proceeding with a number of conclusions, it is necessary to make three remarks.

- Almost all resolution measures which are described in this report are calculated from the MTF of the imaging system. It is assumed that this exists. This is in fact highly improbable because it is known that neither the photographical nor the electro-optical imaging system is linear. In both systems the relationship between the luminance of the scene L' and that of the corresponding part of the image L is described in a first approximation by the exponent gamma: $L \propto L'^{\gamma}$. It is rarely the case that γ equals 1.
  It is of course always possible to describe the imaging system in a first-order approximation with the aid of MTFs. It is surprising that this first-order approximation produces such high correlations (up to r=0.98!). Thus one wonders, as did Snyder (30), whether a possible improvement to the perceptual description should not principally be sought in an improvement of the physical (photometric) description of the imaging system.

- Some of the resolution measures described (SMT and CMT acutance, ICS, PLS, VC, SNC) make use of the visual contrast sensitivity, as if it were the MTF of a final additional display system. In doing so the fact is often ignored that there are a number of independent frequency-selective channels in the eye. As a result the eye cannot just be described as a simple band filter. Moreover, the contrast-sensitivity curve is constructed from threshold measurements, whereas it is used for the description of suprathreshold phenomena. This extrapolation is accordingly incorrect. In comparative experiments the course of the "suprathreshold sensitivity" curves is measured (see for example Watanabe (34)). This shows that as the contrast increases, the "sensitivity curve" flattens out and starts to be closer to a low-bandpass characteristic curve. In that sense the shape of the curve in figure 5 is not beyond the bounds of possibility.

- The objective of the comparative experiments described was to find the measure of resolution with the highest possible correlation with subjective sharpness (or quality). It is not absolutely clear what the relationship is between measure of resolution and sharpness: it may be linear, quadratic, logarithmic or otherwise. As far as the **usefulness** of the measure of resolution is concerned it in principle makes no difference what the nature of this relationship is. This correlation is however often calculated on the assumption of a linear relationship (Pearson's product). This is essentially done for reasons of convenience and the lack of a better measure for correlation. Some authors have attempted to improve on this, for example by also calculating the linear correlation with the logarithm of the measure of resolution (Task, section 3.3), or by drawing graphs for all measures of resolution under consideration (Higgins, section 3.2). The correction for the spread of the resolution measures, as applied by Beaton (section 3.4), is the only aspect which **is** taken into account in Pearson's product correlation.

  The more assumptions which have to be made on the anticipated relationship between measure of resolution and sharpness, the greater the chance that resolution measures will be classed as not useful for the wrong reasons. This is important regardless of the manner in which subjective sharpness is to be expressed: thus both in the case of subjective judgments of quality and - perhaps to an even greater extent - in the case of performance measures.

4.2 Unanswered questions

Despite all the work which has been done over the course of the years in the field of perceptual resolution measures, there remains a number of unanswered questions. The most important being:

- Is there a difference between sharpness judgments in task-oriented and non-task-oriented environments?

No experiments have been done in which the same measures of resolution were compared in both types of environment. Nor does experiment II by Synder (section 3.1) really fall outside the task-oriented environment.
It appears as if (see Task's experiments) in task-oriented environments the very low spatial frequencies are not important. On the other hand, the experiment by Higgins shows that in a non-task-oriented environment it is precisely these frequencies which are important: the SQF measure which neglects those frequencies below 10 periods/mm on the retina comes off worse than the comparable measure CMT_{log}, which does take account of the lower spatial frequencies.
Where almost all authors are in agreement is that the higher spatial retina frequencies are visually less important than those in the centre regions. This also appears to be confirmed in the comparative experiments by Beaton through the low score of the SSF measure (section 3.4), which favours the higher frequencies.
The differences found may be attributable to a difference in, for example, experiment configuration. However, it is in principle also conceivable that in a task-oriented environment the lower spatial frequencies are less important than when aesthetic qualities play a role. In other respects too it may be possible that differences between the two environments play a role.

. What is the influence of the scene content?
This question is particularly interesting from the viewpoint of perception. A number of resolution measures contain spectral information of the scene S(u,v) in their original form, or are changed in this respect in the comparative experiments by Beaton. This scene content is incorporated in the resolution measure on the basis of the consideration that it is precisely this scene content which the subject observes. It is not however necessarily the case that this is also the criterion on which the test subject makes his judgment. It is possible that the subject wants to abstract his judgment from the scene content and that up to a certain level he also manages to do so. The comparative experiment 2 by Snyder seems to point in this direction: the number of errors (performance measure) is much more sensitive to scene information than the quality

judgment (categorical scaling) of the subject. It should also be possible to explain in this manner the fact that the correlations in the experiment by Beaton are much greater when they are calculated per scene (type II).

- What is the influence of the image size?
  All comparative studies were carried out with a fixed image format. As a result it is as yet not known what influence the size of the image has. A doubling of the width of the image with fixed system MTF makes it possible to transmit double the quantity of information. Whether this is experienced by subjects as an increase in sharpness or appeals to another psychological dimension is unknown. It does however depend on this whether or not the width of the image must be included in the measure of resolution, as so far only Cohen has done in his Visual Capacity (paragraph 2.2.6).

- What is the influence of noise?
  A number of resolution measures takes account of the noise in the imaging system (MTFA, SNC, JND model). In the MTFA measure the very important threshold curve $m_{d,v+s}(w)$ is determined by this; the two other measures explicitly contain a noise term in their formula. The sole comparative experiment in which the influence of noise is studied is the last experiment by Higgins (section 3.2). However, because it is not clear whether subjective quality or sharpness was the scaling criterion, it cannot be concluded from this experiment whether the noise influences the sensation of sharpness or only the quality. In this latter case, in the same way as the problem with the image size, the noise need not be included in the resolution measure.

4.3 Conclusions

The literature study for a perceptually relevant measure of resolution produced not only a large number of resolution measures also threw up many question marks (see section 4.2). Conclusions can accordingly only be drawn for a very limited collection of situations: namely when pictures with differing resolution, but with fixed dimensions and a fixed viewing distance are compared. Moreover, the drawing of conclusions is made all the more difficult by the fact that almost all authors used the concepts of "quality" and "sharpness" indiscriminately. The following conclusions should therefore also be seen in the light of these restrictions.

The body of the proposed measures of resolution appears on the one hand to be very extensive, but on the other hand to show considerable similarity. The major part essentially consists of an integration of the MTF of the imaging system. This MTF is weighted with information about the eye and through either the division by or the subtraction of the contrast modulation threshold of the visual system. What these two operations have in common is that they consider modulation to be less important at the higher spatial frequencies than that in the middle regions. They differ however in the treatment of the lower spatial frequencies. Further modifications of the resolution measure are then applied through transformations of the spatial frequency axis (for example logarithmic integration) or of the modulation axis (through for example squaring or extraction of the root). As the final modification, the integration limits can also be adjusted.

Strangely enough, the reason for these operations is seldom given. Exceptions are perhaps the TQF of Charman and Olin and the SQRI of Barten (section 2.2.3 and 2.2.9), where at least an attempt is made to do so. Only the JND model of Carlson and Cohen (paragraph 2.2.8) rests on a number of sound model bases. This however results immediately in a complexity which gives rise to extensive calculations, perhaps slightly overdone, and which is in any case unwieldy for a simple measure of resolution.

From the scarce comparative experiments the MTFA, or a measure derived from it, always emerges as one of the best. The correlations found are often as high as 0.92 and never lower than 0.84. That fact is probably the justification for the use of the MTFA in evaluations of imaging equipment, as is already done in the literature, for example by Infante (19). Some caution is however needed here: the three comparative experiments in question which had the MTFA as the final outcome, were carried out in the same sort of environment and all relate to military recognition.

Some resolution measures were never included in a comparative study (SQRI, VC, BLS), or only in modified form (ICS). These measures are closely related to the $MTFA_{log-log}$, which is also fairly satisfactory as a measure of resolution. There is therefore a real possibility that one of these resolution measures could surpass the MTFA.

In conclusion it can be stated that it is still too early to recommend a single perceptual measure of resolution. Answers to the questions of section 4.2 are necessary to provide insight into the conditions under which the required measure of resolution will be valid. A good and unambiguous specification of the visual system will then also be necessary: a curve of contrast sensitivity or of the contrast modulation threshold as a function of spatial frequency. This must be measured under relevant conditions and can possibly have properties of the imaging system as parameter (for example luminance, noise etc.).

The high correlations found between various measures of resolution and sharpness (or subjective quality) indicate however that it must be possible to find and apply such a measure for perceptually weighted spatial resolution.

References

1 P.G.J. Barten, Evaluation of visible resolution on a display, Japan Display, Tokyo, SID/ITE, Sept 30 - Oct 2, 1986. 530-533.

2 P.G.J. Barten, Resolution of projection T.V. systems, SID Digest 17(1986)455-458.

3 R.J. Beaton, Quantitative models of image quality. Proceedings of the Human Factors Society, 27th annual meeting, 1983, 41-45.

4 F.J.J. Blommaert, Vision of Details in Space and Time. Thesis Eindhoven University of Technology, 1987.

5 R. Bracewell, The Fourier Transform and its Applications, McGraw-Hill, New York, 1965.

6 C.R. Carlson and R.W. Cohen. A model for predicting the just noticeable difference in image structure as a function of display modulation transfer. SID Digest (1978)30-31.

7 C.R. Carlson, Thresholds for perceived image sharpness, Photographic Science and Engineering 22(1978)69-71.

8 C.R. Carlson and J.R. Bergen, Perceptual considerations for high-definition television systems SMPTE J 93(1984)1121-1126.

9 W.N. Charman and A. Olin, Image quality criteria for aerial camera systems. Photographic Science and Engineering 9(1965)385-397.

10 R.W. Cohen and I. Gorog, Visual capacity - an image quality descriptor for display evaluation, Proc. SID 15(1974)53-62.

11 E.M. Crane, An objective method for rating picture sharpness: SMT Acutance, J SMPTE 73(1964)643-647.

12 M. Davidson, JOSA 58(1968)1300.

13 H. Davson, The eye 2A, Academic Press, New York, 1976.

14 R.G. Gendron, An objective method for rating picture sharpness: CMT Acutance, J SMPTE 82(1973)1009-1012.

15 E.M. Granger and K.N. Cupery, An optical merit function (SQF), which correlates with subjective image judgments, Photographic Science and Engineering 16(1972)221-230.

16 J.P. Guildford and B. Fruchter, Fundamental Statistics in Psychology and Education, McGraw-Hill, New York, 1978.

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

18 B.R. Hunt and G.F. Sera, Power law stimulus-response models of image quality in nonperformance environments. IEEE Trans. Systems Man and Cybernetics SMC8(1978)781-791.

19 C. Infante, On the resolution of raster-scanned CRT displays, Proc. SID 26(1985)23-36.

20 E.H. Linfoot, Fourier Methods in Optical Image Evaluation, The Focal Press, London, 1964.

21 E.M. Lowry and J.J. DePalma, Sine-wave response of the visual system. I. The Mach phenomenon, JOSA 51(1961)740-746.

22 A. van Meeteren, Visual Aspects of Image Intensification, dissertation, Rijksuniversiteit Utrecht, 1973.

23 C.N. Nelson, Photographic system as a communication channel, Applied Optics 11(1972)87-92.

24 M.D. Sacks J. Nachmias and J.G.Robson, Spatial frequency channels in human vision JOSA 61(1971))1176.

25 O.H. Schade, Image gradation, graininess and sharpness in television and motion picture systems. Part III: The grain structure of television images. J. SMPTE 61(1953)97-164.

26 O.H. Schade, Some characteristics of the visual system and effect on overall (color TV) system concept, J SMPTE 67(1958)813-819.

27 O.H. Schade, J SMPTE 73(1964)81.

28 C.E. Shannon Bell Syst. Tech. J. 27(1948)379+623.

29 C.E. Shannon, Proc. IRE 37(1949)10.

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

31 S.S. Stevens, The Psychophysics of sensory functions, Amer. Scientist 48(1960)226-253.

32 C.F. Stromeyer and B. Julesz, Spatial frequency masking in vision: critical band and spread of masking, JOSA 62(1972)1221-1232.

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

34 A. Watanabe, T. Mori, S. Nagata and K. Hiwatashi, Spatial sine-wave responses of the human visual system. Vision Research 8(1968)1245-1263.

35 A.B. Watson, H.B. Barlow and J.G. Robson, What does the eye see best?, Nature 301(1983)419-422.

36 R.N. Wolfe, Width of the human visual spread function as determined psychometrically, JOSA 52(1962)460-469.