# Literature survey

*Citation for published version (APA):*
van Overveld, W. M. C. J. (1995). *Literature survey: perceived quality of fluoroscopic images*. (IPO rapport; Vol. 1065). Instituut voor Perceptie Onderzoek (IPO).

*Document status and date:*
Published: 14/08/1995

*Document Version:*
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

*Please check the document version of this publication:*

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

*General rights*
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

*Take down policy*
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 17. Nov. 2023

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

14.08.1995

Rapport no. 1065

# Literature survey: perceived quality of fluoroscopic images

W.M.C.J. van Overveld

# Literature survey:
# perceived quality of fluoroscopic images

**W.M.C.J. van Overveld**

# Index

# Summary

This literature survey is a deliverable of the IPO / Philips Medical Systems project "Optimization of the perceived quality of fluoroscopic images". It is a survey of both classical and recent publications about various aspects of image perception. Some aspects of visual perception that have received special attention in this report are those directly concerned with noise-reduced fluoroscopy sequences: the perception of noisy images, perception of moving images (motion blur), and the methods that can be used to evaluate the quality of medical images.

# 1 Introduction

This document a deliverable of the IPO / Philips Medical Systems project "Optimization of the perceived quality of fluoroscopic images". It contains a compilation of references to literature about human visual perception. Various topics are addressed, such the classical experimental work on contrast sensitivity, models for detection of signals in noise, image quality models, and evaluation methods.

Since it is obviously impossible to compile an exhaustive list of the literature in this wide area of research, we only tried to review some of the most important (or the most easily accessible) papers and textbooks in the field. This should allow the interested reader to find more specialized information on the subject. We concentrated on topics that are closely related to the subject of the research project: fluoroscopy, dynamic images in general, perceptual effects of noise, contrast and blur on the perceived image quality, and the ways to measure image quality (or aspects of it) for medical applications. However, since this survey is also meant to be used for future work on image quality, more remote topics have been addressed as well (albeit less thoroughly).

The set-up of this report is as follows. We organized the literature into themes like "the human visual system in general", "perceptual effects of noise", "temporal effects" and "evaluation methods". The following chapters each address one of these themes and briefly discuss the literature dealing with that theme. For some important papers we include a summary of the results, but other papers are mentioned only in passing. In the chapters, we refer to the papers as [ABC*yy*] where A, B and C are the first letters of the authors' surnames (followed by + in case of more than four authors) and *yy* refers to the year of publication. In case of just one author, the first three letters of his name are used: [Aut*yy*]. The full references can be found in the list at the end of the report. For those interested, almost all papers and books are available through the author of this report.

Due to the organization in themes, it happens that some papers are mentioned in more than one chapter. As an example, [RMH95] studies the effect of background complexity ('anatomical noise') through an unconventional evaluation method, namely the reaction time needed to locate the target. Since noise, including anatomical noise, is treated in Chapter 3 and evaluation methods are treated in Chapter 5, the paper occurs in both chapters.

# 2 The human visual system: some properties and their consequences

In this chapter we briefly describe some of the "classical" literature on the human visual system (HVS): perception of luminance, luminance differences, contrast differences, and some effects of noise. Some more application-oriented papers are discussed as well.

General textbooks about human visual perception are e.g. [Gra68], [Cor70] and [Gra65]. These treat psychophysical findings as well as physiological data. Physiology is treated extensively in [Dav64] but will not be discussed further in this report. The early psychophysical work is mostly empirical. Especially detection and discrimination behaviour (the minimum luminance or colour difference that can be seen, or the shortest light pulse...) have been studied extensively. From this work, various models have been developed which (partly) explain how luminance patterns are processed in the eye and brain.

Since the late 1970's, there is an increasing interest in a field called computational vision. Here, the main topic is image understanding or "inverse optics": how is a 2-dimensional intensity image represented and transformed in early vision processes and how are physical properties of the real world (e.g. edges and curvature of surfaces) recovered from the image? A good example of a computational theory is given in [Mar82]. This theory distinguishes two stages: the intensity image is transformed to a "primal sketch" (using zero crossings of second derivatives at various scales to detect intensity changes and local geometrical structure) and this is transformed to a "$2^1/2$-D sketch" in which the geometry of visible surfaces is represented. A review of this type of work can be found in [PTK92]. For this and other recent topics in the study of visual perception, we refer to [Bla90].

Other useful review papers about properties of the HVS and especially about how these properties are exploited in medical imaging are [PH91], [Kun86], [Che82] and [Che92]. [PH91] describes a model of the HVS, similar to the one in [Mar82]. The model is used to explain phenomena such as Mach bands, Weber's law, masking and adaptation. The paper mentions some implications of this for the presentation of medical images (ambient light, use of pseudocolour,...). In [Kun86], less attention is paid to what the visual system actually does, but instead the consequences for medical images are discussed in a comprehensive way. The paper also stresses the importance of appreciation-oriented quality aspects: a point which is overlooked in most other publications. The first sections of [Che82] and [Che92] also describe the HVS, both in terms of physiology and in terms of performance measures (modulation transfer, threshold contrast,...). The remainder of [Che92] concentrates on detection, w.r.t. measurements (ROC, treated in section 5.1 of this report) and models (Statistical decision theory; see section 3.1). The same holds for [Che82], although this paper also discusses the usefulness of various image enhancement techniques.

To come back to the "classical" detection work: W.r.t. the perception of luminance differences, i.e., the detection of an object with luminance $L+\Delta L$ on a background of luminance $L$, [Bla72] gives a good review starting from Weber's work and concentrating on contrast-detail studies (i.e., investigating how $\Delta L$ depends on the size of the object to be detected). For a given object size, the threshold contrast $\Delta L/L$ decreases with $L$ for low values of $L$ ("Rose - de Vries region"), and is constant for high $L$ ("Weber region"). Various models have been proposed (cf. [SYD87]) to describe this and to construct a mapping from luminance to perceived brightness from the data. For recent work on the luminance-brightness mapping, see [Sch93]. The threshold contrast as measured for sinusoidal gratings depends on both frequency and luminance, as reported in [NB67]. The minimum threshold lies at 1 cyc/deg for low luminances (0.1 cd/m$^2$) and at 4 cyc/deg for high levels (100 cd/m$^2$).

Luminance discrimination has also been studied. In this case, two objects with luminances $L_o$ and $L_o + \Delta L$ are both superimposed on a background of luminance $L_b$. The minimum $\Delta L$ needed for discrimination depends on $L_o$ and $L_b$, but also on the ambient light, as shown in [RJP87]. Ambient light affected both physical contrast and the viewer's adaptation level, but the adaptation level had little effect on the discrimination thresholds.

Going one step further, we arrive at contrast discrimination: discriminating between a pair of objects (or a sinusoidal grating) with contrast $C$ and a pair with contrast $C+\Delta C$. This has been studied in [LKB87], [NS74], [LF80]. It was shown that with a low "pedestal" contrast ($C$), discrimination thresholds are lower than detection ($C$=0) thresholds so that the visual system is more sensitive in the presence of the pedestal. This is a "facilitation" effect. With larger pedestals, however, the discrimination threshold increases, which is called a "masking" effect. Some models have been developed to describe these effects.

Contrast sensitivity is also affected by noise. The following papers treat the topic of detection of sinusoidal gratings in the presence of noise. Also see section 3.1. Pollehn and Roehrig [PR70] measured the minimum contrast needed for the visibility of sinusoidal patterns, when white noise or $1/f$ noise was added. The results could be modelled by

$$S_t = a\sqrt{N^2 + N_i^2}$$

in which $S_t$ is the signal at threshold, $N$ is the noise standard deviation and $N_i$ is internal visual noise. This model has been applied more widely (cf. [Dal93]).The authors also found that for low noise, the signal-to-noise ratio (SNR) at threshold ($a$ in the formula) was minimum for a frequency of 7 cyc/deg, but the minimum shifted to 2 cyc/deg with increasing noise. Stromeyer and Julesz [SJ72] also considered the masking effect of noise. They measured contrast thresholds of sinusoidal gratings for dynamic 1-dimensional filtered noise. The effect was measured in terms of threshold elevation (defined as the threshold for the noisy case divided by the threshold without noise, minus 1). The threshold elevation was maximum for grating frequencies close to the centre frequency of the band-limited noise, and it decreased by half when the two frequencies were 0.5 to 0.75 octaves apart. The threshold elevation increased when the bandwidth increased from 0.5 to 1 octave, but wider bands had no further effect even though the noise power increased.

From the data of [SJ72], Barten [Bar95] derived a model for spatial frequency masking and the masking effect of pedestals in contrast discrimination. The model was checked on data by Van Meeteren and Valeton ([MV88]; chapter 34 of [Bla90]), who measured contrast sensitivity in the presence of filtered 2-dimensional noise. They found that the threshold contrast depended on the grating frequency together with spectral composition of the noise: coarse noise (bandwidth 4.5 cyc/deg) raised the threshold for low grating frequencies (1-4 cyc/deg) but fine noise affected the threshold for all frequencies.

# 3 Perceptual effects of noise

Since this report was written in the context of the "optimization of the perceived quality of fluoroscopic images" project and this project is largely concerned with noise reduction, we devote this chapter to the perceptual effects of noise. There are two separate effects which we must consider, namely a) the masking effect of noise (the fact that signals, or relevant image data are less easy to see in the presence of noise) and b) the visibility and annoyance of noise itself. Effect a), which has been studied much more extensively than effect b), is mostly related to performance-oriented or diagnostic image quality. The literature relating to this is described in section 3.1. Effect b), on the other hand, has to do with appreciation-oriented or cosmetic quality. This is considered in section 3.2.

## 3.1 Visibility of signals in the presence of noise

Early work on detection of objects when detection is limited by quantum noise is that by Rose (1942, 1948) and De Vries (1943). A comprehensive overview of this classical work is given in [Ros73], and a more general review including extensions of the model can be found in [Sch73] and in section 4.1 of [Che92]. Rose proposes the "ideal photon counter model" to predict the necessary contrast (in number of photons) of a disk on a homogeneous background. If the background contains $N_b$ quanta and the contrast $C$ is $\Delta N/N_b$, the model predicts a detectability in 50% of the cases if the SNR $k$, defined as $CN_b/\sigma_b$, equals 5. With a Poisson model for the distribution of quanta, $\sigma_b$ equals $\sqrt{N_b}$. The number of quanta can be written in terms of the number of quanta per unit area per second, the diameter of the disk, and the sampling time (effective eye integration time).

For a specific medical application of the Rose model, see [RKY95]. It is used for the detection of disks of a given diameter D and thickness t in an X-ray phantom. Here, the SNR at threshold is given in terms of x-ray photon fluence, attenuation of background tissue at a given energy, the DQE (detective quantum efficiency) for disks with diameter D at energy E, the X-ray contrast of the disk (the linear x-ray attenuation at energy E times the thickness), and the scatter-to-primary ratio. The authors have applied this to physchophysical measurements in a contrast-detail study and found a fair agreement between experimental data and the model.

The Rose model is extended in [HC76], [Che82] and [CH83]. For small diameters of the disk to be detected, the model is corrected for by taking into account the point spread function due to the HVS, and adapting the sampling aperture used in the Rose model to this. For large diameters, lateral inhibition plays a role. The extended model describes this mechanism as well. This model successfully predicts the results of contrast-detail experiments, as reported in [Che82] and [CH83]; it shows that the threshold contrast is proportional to the external noise power. More examples of the use of this model and related models can be found in [OCD+86], [GD87], [MRBW90], [Bar90a].

The model has also been extended using statistical decision theory. This theory allows one to define an "ideal observer" operating as a maximum likelihood detector, which means that the observer cross-correlates the image with the expected signal. A "quasi-ideal observer" can also be defined similar to an ideal observer without taking into account noise correlations (prewhitening the noise). Human performance can then be related to the performance of this ideal observer. One of the main papers describing this theory in great detail is [WB85]. Many more papers exist on this topic (including some of the papers mentioned in the previous paragraph), but some examples are mentioned below.

In [BWJ82], the authors present a model predicting the detectability of objects in a noisy background. The SNR, computed from the spectral composition of signal and noise, serves as a prediction for the detectability index *d'* as measured in 2-alternative forced-choice (2AFC) experiments. Human performance is compared with the (quasi-)ideal detector. Results are shown for detection of bars with fuzzy and sharp edges, in noisy and noiseless backgrounds. Using the model, the authors also investigate the possible improvement in detection that can be reached by contrast stretching or noise reduction. They argue that contrast enhancement may be more beneficial than noise reduction because when image noise is reduced, internal noise will play a larger role.

[LKB87] describes experiments on discrimination of contrast $C$ vs. contrast $C+\Delta C$ in the presence of noise. Both gratings masked by dynamic noise and disks masked by static noise were studied. The effect of the pedestal contrast $C$ was also studied (cf. chapter 2). The discrimination threshold was modelled as $E_t = k(N+N_{eq})$. Here $N$ is the noise power and $E_t$ is the signal energy at threshold. $N_{eq}$ is called the equivalent (intrinsic) noise. Sampling efficiency is defined as $E_t/(Nk)$. It was shown that $N_{eq}$ increased with the pedestal value, but the sampling efficiency was hardly affected by the pedestal. It was also shown that dynamic noise gave rise to lower sampling efficiency than static noise. Equivalent noise models and related work on human observer efficiency are also reviewed in chapters 1 and 38 of [Bla90].

The above mentioned papers all deal with X-ray quantum noise. Other types of noise have also been investigated, e.g. structure mottle (cf. [KDOG86]), camera noise (cf. [Luij94]) and quantization noise (cf. [Bur85]). A different kind of noise is the so-called structured (or anatomical) noise, which was investigated in the following papers.

The first papers in which an attempt is made to quantify the effect of structured noise are [RKG74] and [KR76]. The authors defined local "complexity" in the vicinity of a lesion as the mean Laplacian (density change tangential to the contour of the lesion). The "conspicuity" of the lesion was then defined as the mean contrast of the lesion with its background, divided by the complexity. In an experiment with clinical chest images and superimposed nodules, the percent correctly detected nodules increased linearly with log(conspicuity) up to a maximum conspicuity value for which detection was 100%. The maximum conspicuity value seemed to depend on the task.

Detection of stenoses in vessels was studied in [ODMG88]. Simulated images of cylindrical vessels with and without lesions were superimposed on uniform quantum noise images of varying dose levels. Subjects had to locate the lesion in an 18AFC experiment. Parameters that were varied: lesion size and location (edge or centre), presence or absence of background structure (i.e., the vessels), vessel diameter, and noise level. It was found that threshold contrast was larger when structured noise was present. The threshold also increased with noise, with decreasing lesion size and with decreasing vessel width.

In [RMH95] the 'saliency' of lesions in simulated angiograms was measured as a function of the local curvature of the blood vessel. The curvature was used as a means to quantify the complexity and it was shown that stenoses located at points of high curvature were found less easily than stenoses at average or low curvature locations.

In [BVHV95], the authors studied whether the ideal detector (cross-correlator) model, used to predict detectability of signals in the presence of system noise, is useful to describe the effect of "anatomical noise" (anatomy plus system noise) as well. This turned out not to be the case because phase information (not used in the cross-correlator model) was shown to play an important role in anatomical noise. Also, for mammographic glandular tissue, anatomical and system noise did not give rise to different detectabilities in a 2AFC experiment, so that the viewers appeared to be able to abstract from the anatomy. A similar failure of the cross-correlator model was found for Gabor patterns embedded in natural (non-medical) images (cf. [CM86]). An "indirect" cross-correlation based on spatial decomposition of the signal and signal+background gave a better prediction.

## 3.2 Visibility of noise in the presence of signals

One of the first papers on the subjective impression of noise is [Hua65]. This paper addresses the annoyance of additive Gaussian noise with different bandwidths in *x*- and *y* direction, for TV-like black-and-white still images. The author found iso-preference curves in the *kx*,*ky* plane (where *kx* and *ky* are the cut-off frequencies of the noise power spectrum in the *x* and *y* direction) with a saddle point, the location of which depended on the scene. Noise containing frequencies similar to those of the original image was found to be less annoying than other types of noise. For equal noise power in both directions, the maximum annoyance occurred around $kx = ky = 0.2$ cyc/min of arc.

Marmolin and Carlström [MC85] investigated the visibility of noise for static gaussian white noise on uniform images of different luminances (23 to 64 cd/m$^2$ on a monitor with $\gamma=1$). They found that noise of a given standard deviation (rms) decreased in visibility when luminance increased, but noise rms divided by average luminance was constant for a fixed impression of noise. They also studied visibility of noise for gradual and stepwise changes in image intensity, but they did not find a clear masking effect.

Girod [Gir89] reports on the visibility of noise under various circumstances: noise located near a spatial or temporal edge, and as a function of background luminance. He presents a model for the HVS and uses it for the prediction of the visibility of coding artefacts (also see section 4.1). The model was tested with experiments on the visibility of noise and gave a good prediction. Girod found that a thin band of dynamic white noise was less visible within ~5 arcmin of a high-contrast edge, especially on the dark side of the edge. For temporal edges, a 40 ms dynamic white noise flash was harder to see within ca. 40 ms after a big temporal brightness jump. Low-contrast edges in the spatial or temporal domain, however, could facilitate detection of artefacts. The variance of just visible noise decreased sharply when the uniform background grey value increased from 0 to 50; noise was best visible around grey value 83, after which the threshold slowly increased again. These values hold for a monitor with $\gamma=2.2$ and maximum luminance 78 cd/m$^2$. These luminances are lower than those used in [MC85], which may explain the different results.

In [Kay95], Kayargadde measured perceived noisiness and modelled the data using a noise estimation algorithm (based on uniform regions, detected using a polynomial transform of the image). He showed that for gaussian, triangular or uniform noise, the noisiness in a uniform image depended only on the standard deviation of the noise and not on the shape of the distribution. Noisiness was also independent of the luminance of the region, provided it was surrounded by other regions such that the whole image had a constant average luminance (note the difference with [MC85] and [Gir89]!). For natural images, Kayargadde defined $\eta$ as $\sigma_n/l_c$ where $l_c$ is the correlation length of the filtered noise and $\sigma_n$ is the standard deviation of the unfiltered noise. For constant $\eta$, noisiness increased with $\sigma_n$ (and $l_c$). For $l_c < 1.6$ arcmin, noisiness was linear with $l_c$ and at higher values, it saturated. Noisiness also increased with $\eta$.

# 4 Image quality: experiments and models

Many researchers have investigated how perceived image quality can be predicted from physical properties of the display system or from the image itself. Since image quality in general is a complex notion, most of the early work (and some recent work as well) concentrated on the visibility of certain details in images. Thus a performance measure was used as a measure for the image quality. Especially in the medical field, this is a widely accepted way to look at image quality. Image compression is another area in which visibility of details is a useful criterion for image quality: the better visible the differences between an original and a compressed image are, the lower the quality of the compressed image will be. Image quality measures based on detectability are described in section 4.1. For a more general notion of image quality, it is necessary to look at several aspects of the perceived quality, like global contrast, sharpness and noisiness. These aspects have been studied experimentally, and attempts have been made to predict the perceptual attributes from physical parameters of the display or the image. This work is treated in section 4.2. Section 4.3 mentions some research on the way how different aspects of image quality are combined into one overall image quality judgement. The section describes empirical work and some mathematical models.

A nice overview of different types of image quality models is given in [Lub93]. In this paper, four levels of models are distinguished: model-free data collection, task modelling, performance modelling, and mechanistic modelling. Model-free data collection corresponds to some of the empirical work described in section 4.2 and 4.3: varying parameters and evaluating the effect. Task modelling is used when a simple detection task can be studied in isolation (e.g., quantization effects: is a small luminance difference visible?). Performance models are based on just noticeable differences and are thus described in section 4.1. Performance models do not try to describe how the visual system actually works, but treat it like a "black box" Examples are [Bar90b], [RC73] and [CC80]. This is opposed to mechanistic models, which give a mathematical description of several stages of the HVS as described in chapter 2. Examples: [Dal93], [Gir89], [Wil91]. Mechanistic models usually capture three aspects of the visual system in the sense that sensitivity (the inverse of the threshold contrast of a wave form) depends on 1) light level (adaptation effects: amplitude nonlinearity of the system, luminance-to-brightness mapping) 2) spatial frequency (eye optics, contrast sensitivity function) and 3) signal content (masking effects). These models are also described in section 4.1.

## 4.1 Image quality models based on detection data

Early work on the modelling of performance-oriented image quality was done by Rogers and Carel [RC73], who modelled the detection and recognition of army vehicles as a function of stimulus size (i.e. subtense of a sinusoidal grating) and frequency, stimulus and surround luminance, and modulation type (horizontal or both horizontal and vertical). Linear regression was used to describe the modulation at threshold as a quadratic polynomial of the logarithm of the physical parameters.

Barten [Bar90b] introduced a model for image quality (or rather, display quality) called the square-root integral (SQRI) method. Its basic elements are the modulation transfer function (MTF) of the display and the contrast sensitivity function of the HVS which can be seen as the MTF of the eye. The SQRI expresses the display quality in units of just noticeable differences (jnd's). The MTF of the display can be easily calculated, but the MTF of the eye cannot. For this, Barten adopted an approximation formula which includes angular display size, display luminance, and various fitting parameters to fit existing experimental data to the formula. More recently (cf. [Bar91]), the model was enhanced to include noise. Here the approximation function for the MTF of the eye was extended with two further parameters dealing with noise: one for threshold behaviour of the eye, and one for supra-threshold behaviour. Barten's model is a simple case of the model by Carlson and Cohen [CC80], in which separate frequency bands are considered rather than the integral over all frequencies.

Many other models have been developed especially for the evaluation of image processing and image coding algorithms. A good example is the one by Daly [Dal93]. This reference also gives a good review of other models and compares them with Daly's model. Some more examples are given below.

The model of [SCH89] is introduced to facilitate the evaluation of DCT coded images. This model computes the so-called information content (IC) of an image. First "intensity" values are mapped to brightness values using a power law. Then a DCT is performed to find the spectral components of the image and a weighting function is applied for each frequency according to the HVS's contrast sensitivity function. This weighting is defined for different resolutions. The IC for a given resolution is the sum of the magnitudes of the weighted DCT coefficients for that resolution. The IC versus resolution curve can be used to compare the information content of images coded in different ways. The authors do not show how IC is related to the opinion of observers, but they claim that the (many) parameters in their model can be tuned to give a "good" fit to observers' responses.

Another model used for evaluation of coding algorithms is introduced in [Gir89]. This vision model includes the monitor, direction of gaze, and photoreceptor activity. It predicts the perceived difference between an "original" and "impaired" image by analysing how infinitesimal changes in an image are affected by the monitor, by the optical point spread, and by processing in the fovea (inhibition, adaptation and saturation mechanisms). The linearization procedure thus provides weighting factors for the importance of the effects in each stage of the model. The model was verified using experiments on the visibility of noise, as mentioned in section 3.2.

## 4.2 Measuring and modelling perceptual aspects of images

In this section, we describe some of the literature on measurements and models for contrast and sharpness. The perceptual effects related to noise have already been described in section 3.2.

In [Pel90], Peli proposes a definition of local band-limited contrast in images. For a given location in an image and for a given frequency band, the contrast is defined as the intensity in the band-pass filtered image divided by the intensity in a low-pass filtered image. This local contrast takes into account the fact that contrast sensitivity depends on both the spatial frequency and the local background value. It can be used to predict the visibility of details at various spatial scales and to compare the effects of different contrast enhancement techniques.

Lillesæter [Lil93] proposes another definition of contrast in complex images. According to this definition, the contrast between object and background is composed of area contrast (defined as the average $\log(L_{obj})$ minus the average $\log(L_{backgr})$) and edge contrast, defined as $\log(L_{obj}/L_{backgr})$ integrated over the contour separating the object from the background. It is claimed that this definition of contrast reflects the "potential visibility" of the object against the background, but experimental verification has yet to take place.

Roufs et al. [RKT94] modelled the perceived global contrast by a quantity called "effective gamma", which is similar to the ordinary gamma but it takes the image histogram into account. The authors also investigated how the global contrast affected the image quality. Kayargadde [Kay95] refined this model and used the slope of the cumulative histogram of a brightness image as an estimate for the perceived contrast of the image. The link between physical contrast, perceived contrast and perceived quality has been studied for X-ray images in [Ove94].

Sharpness of static images has also been modelled. (For motion induced impressions of sharpness, see section 6.1.) Kayargadde and Martens ([Kay95], [KM94]) developed an algorithm by which the sharpness of edges in an image is estimated from the image itself. Their algorithm detects edges using a polynomial transform and compares the parameters of the edge with those of an ideally sharp edge. The model also uses knowledge of the relationship between physical edge parameters and perceived sharpness (derived from Nijenhuis [Nij93], who modelled perceived sharpness as a function of blur spread). An earlier investigation of the relationship between blur spread and sharpness on the one hand and sharpness and quality on the other hand can be found in [Wes91].

Another approach to modelling sharpness was introduced in [ER94]. The authors present an algorithm which first finds the boundary of an object and then averages the derivatives in the direction normal to the boundary. Comparisons with sharpness as perceived by observers have not been conducted.

## 4.3 Supra-threshold image quality models

A general overview of image quality measurements for different kinds of imagery is given in [Rou92]. Scaling, matching and performance measures (for legibility of text) are treated using examples from image coding, effects of TV system parameters, and assessment of text terminals. As an example of empirical work, Van der Zee and Boesten [ZB80] have investigated the effect of luminance and size on image quality for slides of outdoor scenes and portraits. They showed that quality increased with both average luminance and size. In [RV93], the effect of noise and blur on the quality of natural images was studied. It was shown that blurring gaussian noise before it was added to a (sharp) image made the image look more sharp and more noisy than when the noise had not been blurred, but the overall quality did not change. In the medical field, [Ove94] and [Ove95a] describe measurements of image quality as a function of various physical parameters (the contrast parameters gamma and grey value range, the width of a blurring kernel, and X-ray dose).

Hunt and Sera [HS78] modelled image quality using a power law: the viewer's response $R$ was modelled as $k(s-s_0)^\gamma$ where $s$ and $s_0$ represent the stimulus strength and its threshold value, respectively. In experiments with images degraded by noise and blur, it turned out that $k$ and $s_0$ depended on the SNR.

A lot of work has been done regarding summation rules: the way how different perceptual attributes combine into an overall quality impression. In general, a Minkowski metric is assumed ([Rid92]), which means that the effects of different perceptual dimensions are raised to the *p*th power and the sum of these is raised to the power $1/p$. Some examples: in [Pfe84], $p=1$ was used, whereas the authors of [RV93] and [Nij93] found that $p=2$ gave the best fit to their experimental data. The latter value means that different perceptual effects can be considered as perpendicular vectors in a euclidean space which can be added using ordinary euclidean arithmetic. This is an example of the multidimensionality of image quality, further explained below.

Marmolin and Nyberg [MN75] were the first to describe perceived image quality as a multidimensional concept. By identifying different subjective dimensions, and studying how these dimensions depended on physical properties of the image, image quality was defined as a weighted sum of the dimensions, where the weighting could depend on the viewer or the task for which the image is used. Marmolin and Nyberg developed such a model by varying images w.r.t. MTF, noise level and spectrum, and greylevel-to-luminance characteristic. Subjects graded the quality of these images and the dissimilarity between pairs of images. The dissimilarity data could be modelled in a four-dimensional perceptual space. (For a description of how a multidimensional space can be constructed from rating data, see [GCS89] or [AN93].) For images in non-performance environments, the authors found that the most important dimension could be interpreted as sharpness; the second one as noise, and the third as contrast. The authors also computed some objective measures of image quality, such as MTF-area and SNR. These were correlated with the quality judgements and the quality dimensions.

Multidimensional descriptions have also been used to describe the effects of noise reduction in CT images ([ERM94], [Esc92]) and the simultaneous effects of noise and blur ([Kay95]). In [Kay95], a 2-d perceptual space was found, in which unsharpness and noisiness were not completely independent: the directions formed an angle of 80°. Quality was somewhat closer to unsharpness (33°) than to noisiness. Kayargadde also found that constant noise intervals (in terms of standard deviation $\sigma_n$) were perceived bigger when blur increased; constant blur intervals (blurring kernel width $\sigma_b$) were perceived smaller when noise increased. At low blur levels, unsharpness increased with $\sigma_n$ but at high blur levels, it decreased with $\sigma_n$. At low noise levels, noisiness was independent of $\sigma_b$ but at high noise levels, it increased with $\sigma_b$.

# 5 evaluation methods

In this chapter, we review the different ways in which images can be evaluated. Performance-oriented quality is considered in sections 5.1 and 5.2, where detection of details and other tasks are used to assess quality. Section 5.3 is devoted to scaling techniques and related methods, which can be used to evaluate appreciation-oriented quality.

## 5.1 Detection studies

In detection studies, the goal is to assess how well certain details can be seen with a given imaging system. Different displays or processing algorithms or even imaging modalities can be compared w.r.t. usefulness in this way. The classical way to study detection of details is via a contrast-detail study. Here, many circular objects of varying size and/or contrast are shown to the observer, and the observer has to decide which objects are still visible and which aren't. For recent examples of this technique, see [EC94], [RKY95] or [HWBW95]. Another fast way to measure thresholds of detectability is via the transformed up-down method. This adaptive technique (increase the intensity of a signal if the observer does not see it and decrease the intensity if he does) is well known in psychophysics but little used in the evaluation of medical images. Some exceptions are [KPJ90], in which the method is explained and it is applied to the detection of low-contrast detail in CTs, and [Luij94], in which it is applied to the visibility of noise.

Far more popular, but also far more complex, is the receiver operating characteristic (ROC) approach. Especially for medical applications, this technique can give useful information; not only about the probability of correctly detecting a detail, but also on the errors ("false alarms" and "misses"), depending on the decision criterion. Extensive reviews of the ROC method are given in [Met86] and [Che92]. If localization and identification of the objects are part of the task, different versions of ROC can be used, as reviewed in [SMLG75] and [Swe93]. One of the first studies in which ROC was used on a large scale is reported on in [SPW+79]. In this study the diagnostic accuracy of CT is compared with RN (radionucleide) scanning. The paper gives many practical details of the way the study was conducted. A comparison between the M-alternative forced-choice paradigm (yet another method to measure detection performance, common in psychophysics but not often used for medical applications) and ROC studies was recently published in [Bur95]. It was argued that the one method is more useful than the other depending on the validity of assumptions, practical limitations (number of images), and the required accuracy.

A problem with ROC studies is the image material. When real clinical images are used, it is very difficult to obtain a set of "sufficiently subtle" cases, and it is hard to establish the truth value. Clinical images were used in e.g. [SPW+79] and [MMD+88]. At the other end of the scale, there are the classical detection studies where disks have to be detected against a uniform background. These are described in many of the papers mentioned in section 3.1. A slightly more complex situation is the one where the size of the target disk is not known to the observer. In [JKFS95] it is shown that this uncertainty decreases the detection performance. This observation is discussed more generally in [Pel85]. Modelling of this type of results (for simple stimuli) is discussed in section 3.1.

Several in-between solutions have been used. One can go from the disks on the uniform background to disks on a slightly more complex background ("lumpy background", cf. [RB92] and [MRBW90]). One can also try to simulate clinical images with lesions: spherical lesions in a cylindrical blood vessel [ODMG88] or a more complex simulation as used in [RMH95]. One step further is take actual X-ray images, but to use an antropomorphical phantom (e.g. a chest phantom) in which the abnormalities can be simulated by changing something in the phantom. This was used in [SRN+78]. Finally, real clinical images can be used (known to be normal), and lesions or other details can be superimposed on them by image processing. This technique was used in [Ove95a] and in [PPM+95].

## 5.2 Indicators of performance not using detection tasks

There are a few papers where performance-oriented quality is assessed with techniques differing from detection tasks. The work described in [CGO94] and [PTC+94] is an example of this. Here observers had to measure the diameters of blood vessels in images. It was argued that this task is relevant, because radiologists use the width of the aorta to decide whether or not an aneurism is present, and the therapy depends on this measurement. The authors studied the effect of image compression by comparing the measured diameters for the original (uncompressed image) with those for images compressed at various bit rates.

Speed can also be used as an indicator of performance. This can be the speed of manipulation, as used in [FMPR88]. Here the speed needed to move a catheter tip to a prescribed position in an angiographic model system was used to compare the usefulness of pulsed and continuous fluoroscopy. The time needed to complete a searching task can also be used. This was used to assess the quality of video display units (searching a pseudotext for the occurrence of a given letter, cf. [RB91]). Recently, searching speed was used in a medical context as well. In [RMH95], the speed of finding abnormalities in images (simulated angiograms with simulated stenoses) was used as a measure for detectability or 'saliency'.

Another tool in the evaluation of image quality is eye movement analysis. Saccade (jump) lengths, fixation locations and durations, and pupil dilations can be used as indications of how difficult a task is or where the salient features of an image are. Some examples: in [VE84] and [Val83], the authors describe patterns of eye fixation locations during the reading of thorax images, which were shown to vary with the instructions given to the observer. In [BPD+94] and [BD95], eye movements were studied for mammographic images. In [BPD+94], patterns of fixation durations and changes in pupil size have been correlated with the results of an ROC study. In [BD95], the authors tried to model the pattern of dwell lengths as a Markov chain. Some applications not related to medical images in which eye movements have been used to analyse the image quality: text terminals (cf. [RB91]), for which the saccade length is used as a measure for legibility of text; and TV images (cf. [Hea93]), for which different TV scenes have been analysed to see where the viewer's attention is drawn, so that other (moving) parts of the image may be coded with fewer bits (also see chapter 6).

## 5.3 Assessment of “overall quality”

Scaling techniques are well known in psychophysics. There are some text books treating the topic in general, like [Dun83]. For the quality of medical images, scaling has not been used to great extent.

In [PTC+94] and [CGO94], subjects were asked to assign scores from 1 (worst) to 5 (best) for the “usefulness for the measurement task” for images in which they had to measure the diameter of blood vessels. This definition of usefulness was supposed to reflect the amount of blur or distortion of edges, and the observer’s confidence in the measurement. The usefulness scores were compared to the measured vessel diameters for different compression ratios. The authors found that usefulness scores decreased drastically with the compression ratio, but the measurements remained fairly accurate for all compression ratios. In [CGO94], the authors also report “usefulness” scaling for CT lung images, as compared to the outcome of an ROC study. Again they found a decrease of usefulness scores when the ROC results remained accurate. The authors’ conclusion is that the subjective usefulness scores do not predict the diagnostic quality (utility for the task) of the images very well. This seems to suggest that it is not necessary to ask for usefulness scores, because measurements are more reliable. Note that this is in conflict with the conclusions of [Ove95a], where a similar discrepancy between the results of a task and the subject’s opinion of image quality was found. Here the conclusion was that it was necessary to assess quality in both ways, because both types of quality were relevant. In [Ove95a], the task was a 2AFC task, where the subjects had to detect a blood vessel which could occur in two locations. The parameters that were varied in this case are: grey value range, X-ray dose and the kernel width of a gaussian blurring filter. In the scaling part of the experiment, subjects had to rate the “overall quality” of the image on a scale from 1 to 10. The quality scores varied mostly with the blur parameter, whereas the probability of correct detection depended mainly on the noise (dose) parameter. The quality scaling part of the experiment is the same as in [Ove94] in which the effect of gamma was studied.

The authors of [MMD+88] also used a rating - or rather, ranking - method to evaluate image quality. They compared the diagnostic quality of thorax images presented on hard copy with those presented on soft copy with either normal or reversed gray scale. Viewers had to express their preference for hard or soft copy, and for normal or reversed gray scale video. The preference was considered in terms of both “ease of reading” and “diagnostic superiority”. An ROC study was also done. Here again it was found that the subjective preference did not always agree with the performance measure. There were cases for which most observers preferred one type of display (w.r.t. diagnostic usefulness), but the area under the ROC curve (the diagnostic accuracy) was significantly larger for the other type. This work is similar to [FVHB87], in which the authors compared the “diagnostic quality” of thorax and genitourinary images presented on different display modalities through scaling of the “ease with which a feature could be seen” on a scale from 1 to 6.

Apart from scaling, other techniques to assess appreciation-oriented image quality have been used. These include interviewing techniques (cf. [Ove95b]) and questionnaires (cf. [GGM+94]).

# 6 Temporal effects

## 6.1 Motion blur

This section is devoted to the effect that moving objects seem less sharp. This observation suggests that the resolution requirements for moving objects are lower than for static objects, which might be exploited in image compression or image processing, for instance. However, the following literature points out that the difference in required resolution is much lower than would be expected.

Glenn & Glenn [GG85] studied the perceived sharpness of moving television images. A horizontally moving complex image was presented at two different resolutions (150 or 300 TV lines) and at speeds ranging from 0 to 15 deg/sec. (For comparison: if a fluoroscopy image of 20 cm diameter is viewed from a distance of 1 m, it subtends 11 degrees visual angle) At each speed, subjects had to indicate which of the two stimuli (at 150 or 300 lines) appeared sharper, in a 2AFC experiment. At higher speeds, the perceived difference between the two images became smaller. These experiments focused on sharpness discrimination rather than assessing the sharpness impression of an image by itself.

Westerink and Teunissen ([Wes91], [WT95]), on the other hand, used scaling to assess the perceived sharpness of moving images. One of their experiments was carried out with the subject fixating a given location of the screen. Sharp and horizontally blurred images were shown moving from left to right at speeds varying from 0.9 to 5.3 deg/sec. The perceived sharpness of the unfiltered image decreased with speed over the whole range. The perceived sharpness also decreased with increasing filter width (from 2 to 6 pixels). For the largest filter widths, the increasing speed had little additional effect. In a second experiment, subjects could pursue the moving image across the screen. The unfiltered image appeared sharp for speeds below 10 deg/sec. For speeds in the range 10-40 deg/sec, the sharpness decreased with speed. For filtered images, the sharpness increased with speed in the range 1-25 deg/sec, and for higher speeds the sharpness stabilized at a level slightly below that of the unfiltered image at the same speed. At speeds above 40 deg/sec, the image started to disintegrate (it was seen as separate frames, jumping across the screen).

Eckert and Buchsbaum [EB93] mention that a high spatial resolution is required when observers are able to track a moving object perfectly. During acceleration of objects, however, some loss of resolution is permissible, especially when this acceleration is not predictable. The paper gives a formula for retinal velocity as a function of the acceleration of an object and relates the minimum spatial screen resolution to this retinal velocity.

## 6.2 Appearance of pulsed fluoroscopy

Pulsed fluoroscopy can suffer from the so-called "Chaplin effect": jerky appearance of moving object due to the low refresh rate of the images. This intermittency effect has been studied in [Kao77] for a task in which subjects used a pen to track random waves. Visual feedback consisted of display of the difference between the hand position and the original wave, where the feedback image was refreshed at intervals of 0 (continuous feedback), 200, 400, 600, 800, 1000 and 1500 msec. It was shown that already at a refresh interval of 200 msec (5 frames/sec), subjects made more errors than in the continuous case. The number of errors increased with the time interval.

Several investigators have studied the dose reduction possible with pulsed fluoroscopy compared to continuous fluoroscopy, under the restriction that the noise appearance is the same for both. Within PMSN, Adriaansz [Adr92] and Verhoeven [personal communication, 1995] have conducted experiments with pulsed fluoroscopy sequences at various dose rates and frame rates. In [Adr92], simulated phantom images were used and images of different dose rates and frame rates were shown in split-screen. For a constant dose speed (dose per frame divided by fame rate), images at lower frame rates were perceived as less noisy. Even when the dose speed for the low frame rate was 25% less than the dose speed for the high frame rate, the former was still less noisy.

Aufrichtig et al. [AXT+94] present a study in which they determined the so-called equivalent perception dose for pulsed fluoroscopy at various frame rates, i.e., the dose at which disks of a given size and contrast are "equally visible" in pulsed and continuous fluoroscopy. The visibility of disks was measured in three ways: paired comparison of a computer generated contrast-detail phantom image, with pulsed and continuous images shown in split-screen; finding the disk of lowest contrast which is still visible, in the same split-screen image; and a 4AFC task in which the observer had to indicate in which quarter of the image a disk appeared. The three methods gave comparable results and showed that, relative to continuous fluoroscopy at 30 frames/sec, a dose saving of 22%, 38% and 49% is possible for frame rates of 15, 10 and 7.5 frames/sec, respectively. A similar study was done by Verhoeven of PMSN, using a matching paradigm for real fluoroscopy sequences displayed at different frame rates.

The usability of pulsed fluoroscopy (at 5 frames/sec) compared to continuous fluoroscopy has also been assessed in [FMPR88], using the speed with which a manipulation task could be completed as an indicator of performance. Few significant differences were found.

## 6.3 Delay

It seems that perceptual aspects of delayed continuous visual feedback in fluoroscopy have not been studied - or at least, not published. All knowledge about delayed feedback comes from fairly old studies in which subjects had to perform tasks like writing, drawing or tracking figures. The subjects could not look at their hands directly, but only through a monitor showing a delayed image. The amount of errors or the time to complete the task was used as a performance measure. Examples of this work are [Smi62], [Kao77], [SB80] and [MM90]. In the experiments described in [Smi62], subjects had to write 2, 4, 8 and 12-letter words with a visual feedback delay of 0, 40, 80, 150, 270 and 520 msec. It was shown that the "neatness" decreased with delay, and, if subjects were instructed to write as neatly as possible, the time to complete the task also increased with delay. Even 40 msec delay was enough to deteriorate the performance compared to the immediate feedback case. Similar results were found in [MM90], where neatness of writing English words and

Kanji letters was affected by a delay of 67 msec and errors occurred frequently at delays of 256 and 500 msec.

In [Kao77], subjects used a pen to track random waves. Visual feedback consisted of continuous display of the difference between the hand position and the original wave, with delays of 0, 200, 400, 600, 800, 1000 and 1500 msec. At 200 msec delay, the subjects already made more errors than at 0 msec, and the number of errors increased with delay.

Smith and Bowen [SB80] describe experiments in which a subject had to move his hand from a given location to a target location, with visual feedback delayed by 66 ms. The subject was trained to move his hand at prescribed speeds of 150, 250, 350, 450, 550, and 650 msec for the distance of 15 cm. Delay resulted in an overshoot (moving too far) at all speeds, but the error was largest (10 mm) around a movement time of 450 msec and dropped to 3 mm at the lowest and highest speeds.

More recently, knowledge about visual feedback has been used in the development of simulator equipment. In [PM92], an overview is given of the requirements for simulators for various types of vehicles. The authors also mention the maximum allowable delay in visual feedback. The maximum delay depends on the reaction speed required to operate the real vehicle. For various ships, delays may vary between 10 and 200 ms. For driving simulations, the delay must not be more than 40 to 80 ms. For flight simulators, 100 to 150 ms is acceptable. Image and motion response should not be shifted in time (at the most, visual feedback may be 150 ms later but not earlier). Helicopter or combat jets allow no more than 40 to 80 ms.

## 6.4 Temporal masking

Stelmach et al. [STW+94] have investigated the sensitivity to coding artefacts following a scene cut. Their work shows that an observer is markedly less sensitive to artefacts in the first frame following a scene cut, but the effect rapidly decreases in subsequent frames. By the third frame, the effect has vanished. This is a similar temporal masking effect as found by Girod [Gir89], mentioned in section 3.2.

# Acknowledgement

Much of the literature mentioned in chapter 5 was gathered in the context of an IPO graduate school seminar "Evaluation of the perceived quality of medical images". I would like to thank Najoua Belaïd for her cooperation in the preparation of this seminar and for the many papers that were included as a result of her suggestions.

# References

[Adr92] M. Adriaansz, "Noise impression related to dose and frame speed", PMS internal memo no. XSB048-221, 1992.

[AN93] A.J. Ahumada and C.H. Null, "Image quality: a multidimensional problem". In: A.B. Watson (ed.), *Digital images and human vision.* MIT Press, London, 1993.

[AXT+94] R. Aufrichtig, P. Xue, C.W. Thomas, G.C. Gilmore and D.L. Wilson, "Perceptual comparison of pulsed and continuous fluoroscopy", Med. Phys. vol. 21, pp. 245-256, 1994.

[Bar90a] H.H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability". J. Opt. Soc. Am. A, vol. 7, no. 7, pp. 1266-1278, 1990.

[BD95] J.R. Barrett and S.J. Dwyer III, "Modeling the visual search patterns of mammographers as a Markov process". Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 116-123, 1995.

[BPD+94] J.R. Barrett, E.S. deParedes, S.J. Dwyer, M.B. Merickel and T.E. Hutchinson, "Unobtrusively tracking mammographers' eye gaze direction and pupil diameter". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception. pp. 46-54, 1994.

[Bar90b] P.G.J. Barten, "Evaluation of subjective image quality with the square-root integral method", J. Opt. Soc. Am. A, vol. 7, no. 10, pp. 2024-2031, 1990.

[Bar91] P.G.J. Barten, "Evaluation of the effect of noise on subjective image quality", Proceedings of the SPIE vol. 1453: Human vision, visual processing, and digital display II, 1991.

[Bar95] P.G.J Barten, "Simple model for spatial frequency masking and contrast discrimination". Proceedings of the SPIE vol. 2411: Human vision, visual processing and digital display VI, 1995 (preprint).

[Bla72] H.R. Blackwell, "Luminance difference thresholds". Chapter 4 in: D. Jameson and L.M. Hurvich (eds.), *Handbook of sensory physiology,* vol. VII/4: visual psychophysics. Springer: Berlin, 1972.

[Bla90] C. Blakemore (ed.), *Vision: coding and efficiency.* Cambridge University Press: Cambridge, 1990.

[BVHV95] F.O. Bochud, F.R. Verdun, C. Hessler and J.F. Valley, "Detectability on radiological images: the influence of 'anatomical noise'". Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 156-165, 1995.

[BWJ82] A.E. Burgess, R.F. Wagner, and R.J. Jennings, "Human signal detection performance for noisy medical images", Proceedings IEEE International Workshop on Physics and Engineering in Medical Images, p. 99-107, 1982.

[Bur85] A. Burgess, "Effect of quantization noise on visual signal detection in noisy images". J. Opt. Soc. Am. A, vol. 2, no. 9, pp. 1424-1428, 1985.

[Bur95] A.E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods". Medical Physics, vol. 22, no. 5, pp. 643-655, 1995.

[CM86] T. Caelli and G. Moraglia, "On the detection of signals embedded in natural scenes". Perception and Psychophysics, vol. 39, no. 2, pp. 87-95, 1986.

[CC80] C.R. Carlson and R. Cohen, "A simple psychophysical model for predicting the visibility of displayed information". Proc. of the Society for Information Display, vol. 21, pp. 229-245, 1980.

[Che82] M. S. Chesters, "Perception and evaluation of medical images". In: P.N.T. Wells (ed.): *Scientific basis of medical imaging*. Churchill Livingstone: Edinburgh - London - Melbourne - New York, 1982.

[Che92] M.S. Chesters, "Human visual perception and ROC methodology in medical imaging". Phys. Med. Biol., vol. 37, no. 7, pp. 1433-1476, 1992.

[CH83] M. S. Chesters and G. A. Hay, "Quantitative relation between detectability and noise power". Phys. Med. Biol., vol. 28, no. 10, pp. 1113-1125, 1983.

[Cor70] T.N. Cornsweet, *Visual perception.* Academic Press: New York - London, 1970.

[CGO94] P.C. Cosman, R.M. Gray, R.A. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy". Proceedings of the IEEE, vol. 82, no. 6, pp. 919-932, 1994.

[Dal93] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity". In: A.B. Watson (ed.), *Digital images and human vision.* MIT Press, London, 1993.

[Dav64] H. Davson, The eye. 4 vols. Academic Press: New York, 1964.

[Dun83] P. Dunn-Rankin, *Scaling methods.* Lawrence Erlbaum Acc., Hillsdale, NJ, 1983.

[EB93] M.P. Eckert and G. Buchsbaum, "The significance of eye movements and image acceleration for coding television image series". In: A.B. Watson (ed.), *Digital images and human vision.* MIT Press, London, 1993.

[EC94] M.P. Eckert and D.P. Chakraborty, "Quantitative analysis of phantom images in mammography". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2167: Image processing, 1994.

[ER94] S.G. Elkadiki and R.M. Rangayyan, "Objective characterization of image acutance". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception, pp. 210-218, 1994.

[Esc92] B. Escalante Ramírez, *Perceptually-assessed digital processing of medical images.* Ph.D. Thesis, Eindhoven University of Technology, 1992.

[ERM94] B. Escalante Ramírez, H. de Ridder and J.B. Martens, "Multidimensional perceptual quality of noise-reduced Computed Tomography images". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception. pp. 119-130, 1994.

[FVHB87] R.A. Fiske, D.J. Valentino, H.K. Huang and H. Blume, "The effect of digital image display format on perceived image quality". Proceedings of SPIE Medical Imaging 1987, SPIE vol. 767, pp. 631-637, 1987.

[FMPR88] S.L. Fritz, S.E. Mirvis, S.O. Pais and S. Roys, "Phantom evaluation of angiographer performance using low frame rate acquisition fluoroscopy". Med. Phys. vol. 15, pp. 600-603, 1988.

[GGM+94]R.C. Gennari, D. Gur, L.M. Miketic, W.L. Campbell, J.H. Oliver and M.B. Plunkett, "Non-portable CR of the chest - radiologists' acceptance". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception. pp. 105-110, 1994.

[GD87] M.L. Giger and K. Doi, "Effect of pixel size on detectability of low-contrast signals in digital radiography". J. Opt. Soc. Am. A, vol. 4, no. 5, pp. 966-975, 1987.

[Gir89] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals". SPIE vol. 1077: Human Vision, Visual Processing, and Digital Display, pp. 178-187, 1989.

[GG85] W.E. Glenn and K.G. Glenn, "Discrimination of sharpness in a televised moving image". Displays, vol. 6, pp. 202-206, 1985.

[Gra65] C.H. Graham (ed.), *Vision and visual perception.* Wiley: New York, 1965.

[Gra68] Y. Le Grand, *Light, color and vision.* Chapman and Hall: London. Second edition, 1968.

[GCS89] P.E. Green, F.J. Carmone, Jr and S.M. Smith, *Multidimensional sclaing - concepts and applications.* Allyn and Bacon: Boston, 1989.

[HC76] G.A. Hay and M.S. Chesters, "Threshold mechanisms in the presence of visible noise". In: G.A. Hay (ed.), *Medical Images: Formation, Perception and Measurement.* Proc. of the 7th L.H. Gray Conference, Leeds, 13-15 April 1976, pp. 208-219.

[Hea93] P.J. Hearty, "Achieving and confirming optimum image quality". In: A.B. Watson (ed.), *Digital images and human vision.* MIT Press, London, 1993.

[Hua65] T.S. Huang, "The subjective effect of two-dimensional pictorial noise". IEEE Trans. Info. Theory, vol. 11, pp. 43-53, 1965.

[HWBW95]W. Huda, J.L. Williams Jr, C.J. Belden and J.L. Williams Sr, "Imaging performance and CR film density". Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 30-39, 1995.

[HS78] B.R. Hunt and G.F. Sera, "Power-law stimulus-response models for measures of image quality in nonperformance environments". IEEE Trans. on Systems, Man, and Cybernetics, vol. SMC-8, no. 11, pp. 781-791, 1978.

[JKFS95] P.F. Judy, M.F. Kijewski, X. Fu and R.G. Swensson, "Observer detection efficiency with target size uncertainty". Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 10-17, 1995.

[Kao77] H.S.R. Kao, "Effects of delay of feedback and intermittency of feedback in compensatory tracking". Perceptual and motor skills, vol. 44 (3, part 2) pp. 1079-1085, 1977.

[Kay95] V. Kayargadde, *Feature extraction for image quality prediction.* Ph.D. Thesis, Eindhoven University of Technology, 1995.

[KM94] V. Kayargadde and J.B. Martens, "Estimation of edge parameters and image blur using polynomial transforms". CVGIP: Graphical models and image processing, vol. 56, no. 6, pp. 442-461, 1994.

[KPJ90] V. Klymenko, S.M. Pizer, and R.E. Johnston, "Visual psychophysics and medical imaging: Nonparametric adaptive method for rapid threshold estimation in sensitivity experiments". IEEE Transactions on Medical Imaging, vol. 9, no. 4, pp. 353-365, 1990.

[KDOG86]Y. Kume, K. Doi, K. Ohara and M.L. Giger, "Investigation of basic imaging properties in digital radiography. 10. Structure mottle of II-TV digital imaging systems". Med. Phys. vol. 13, no. 6, pp. 843-849, 1986.

[Kun86] H.L. Kundel, "Visual perception and image display terminals". Radiologic Clinics of North America, vol. 24, no. 1, pp. 69-78, 1986.

[KR76] H.L. Kundel and G. Revesz, "Lesion conspicuity, structured noise, and film reader error.". American Journal of Roentgenology, vol. 126, pp. 1233-1238, 1976.

[LF80] G.E. Legge and J.M. Foley, "Contrast masking in human vision". J. Opt. Soc. Am. vol. 70, no. 12, pp. 1458-1471, 1980.

[Lil93] O. Lillesæter, "Complex contrast, a definition for structured targets and backgrounds". J. Opt. Soc. Am. A, vol. 10, no. 12, pp. 2453-2457, 1993.

[LKB87] G.E. Legge, D. Kersten and A.E. Burgess, "Contrast discrimination in noise", J. Opt. Soc. Am. A, vol. 4, pp. 391-404, 1987.

[Lub93] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance". In: A.B. Watson (ed.), *Digital images and human vision.* MIT Press, London, 1993.

[Luij94] H. Luijendijk, "Practical experiment on noise perception in noisy images". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception. pp. 2-8, 1994.

[MMD+88]H. MacMahon, C.E. Metz, K. Doi, T. Kim, M.L. Giger and H.P. Chan, "Digital chest radiography: effect on diagnostic accuracy of hard copy, conventional video, and reversed gray scale video formats". Radiology, vol. 168, no. 3, pp. 669-673, 1988.

[MC85] H. Marmolin and A. Carlstrom, "Visibility of noise in homogeneous and textured areas". FOA report C 50010-H2, Swedish National Defence Research Institute, 1985.

[MN75] H. Marmolin and S. Nyberg, "Multidimensional scaling of image quality". FOA report C-30039-H9, Swedish National Defense Research Institute, Stockholm, 1975.

[Mar82] D. Marr, *Vision.* W.H. Freeman and Company: New York, 1982.

[MV88] A. van Meeteren and J. Valeton, "Effects of pictorial noise interfering with visual detection". J. Opt. Soc. Am. A, vol. 5, no. 3, pp. 438-444, 1988.

[Met86] C.E. Metz, "ROC methodology in radiologic imaging". Investigative radiology, vol. 21, pp. 720-733, 1986.

[MM90] Y. Morikiyo and T. Matsushima, "Effects of delayed visual feedback on motor control performance". Perceptual and motor skills, vol. 70, pp. 111-114, 1990.

[MRBW90]K.J. Myers, J.P. Rolland, H.H. Barrett and R.F. Wagner, "Aperture optimization for emission imaging: effect of a spatially varying background". J. Opt. Soc. Am. A, vol. 7, no. 7, pp. 1279-1293, 1990.

[NS74] J. Nachmias and R.V. Sansbury, "Grating contrast: discrimination may be better than detection". Vision Research, vol. 14, pp. 1039-1042, 1974.

[NB67] F.L. Van Nes and M.A. Bouman, "Spatial modulation transfer in the human eye", J. Opt. Soc. Am. vol. 57, pp. 401-406, 1967.

[Nij93] M.R.M. Nijenhuis, *Sampling and interpolation of static images: a perceptual view.* Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1993.

[OCD+86] K. Ohara, H.P. Chan, K. Doi, M.L. Giger and H. Fujita, "Investigation of basic imaging properties in digital radiography. 8. Detection of simulated low-contrast objects in digital subtraction angiographic images". Med. Phys. vol 13, no. 3, pp. 304-311, 1986.

[ODMG88]K. Ohara, K. Doi, C.E. Metz and M.L. Giger, "Investigation of basic imaging properties in digital radiography. 13. Effect of simple structured noise on the detectability of simulated stenotic lesions". Med. Phys. vol 16, no. 1, pp. 14-21, 1988.

[Ove94] W.M.C.J. van Overveld, "The effect of gamma on subjective quality and contrast of X-ray images". Proceedings of SPIE Medical Imaging 1994, SPIE vol. 2166: Image perception. pp. 96-104, 1994.

[Ove95a] W.M.C.J. van Overveld, "Contrast, noise, and blur affect performance and appreciation of digital radiographs". IPO Manuscript no. 1091; to appear in Journal of Digital Imaging, 1995.

[Ove95b] W.M.C.J. van Overveld, "Image quality issues for fluoroscopy assessed through interviews". IPO report no. 1056, 1995.

[PM92] P. Padmos and M.V. Milders, "Quality criteria for simulator images: a literature review". Human Factors, vol. 34, pp. 727-748, 1992.

[Pel90] E. Peli, "Contrast in complex images". J. Opt. Soc. Am. A, vol. 7, no. 10, pp. 2032-2040, 1990.

[Pel85] D.G. Pelli, "Uncertainty explains many aspects of visual contrast and detection discrimination". J. Opt. Soc. Am. A, vol. 2, pp. 1508-1532, 1985.

[PTC+94] S.M. Perlmutter, C.W. Tseng, P.C. Cosman, K.C.P. Li, R.A. Olshen, and R.M. Gray, "Measurement accuracy as a measure of image quality in compressed MR chest scans", Proc. 1st IEEE International Conference on Image Processing, Austin, TX, vol. I, pp. 861-865, 1994.

[Pfe84] U. Pfenninger, "Image quality evaluation by subjective and objective criteria of sharpness and gradation with b/w transparencies". The Journal of Photographic Science, vol. 32, pp. 207-218, 1984.

[PH91] S.M. Pizer and B.M. ter Haar Romeny, "Fundamental properties of medical image perception". Journal of Digital Imaging, vol. 4, no. 1, pp. 1-20, 1991.

[PTK92] T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory". Nature, vol. 317, pp. 314-319, 1992.

[PR70] H. Pollehn and H. Roehrig, "Effect of noise on the modulation transfer function of the visual channel". J. Opt. Soc. Am., vol. 60, no. 6, pp. 842-848, 1970.

[PPM+95] D.T. Puff, E.D. Pisano, K.E. Muller, R.E. Johnston, B.M. Hemminger, C.A. Burbeck, R. McLelland and S.M. Pizer, "A method for determination of optical image enhancement for the detection of mammographic abnormalities". Accepted for Journal of digital imaging, 1995.

[RKG74] G. Revesz, H.L. Kundel and M.A. Graber, "The influence of structured noise on the detection of radiologic abnormalities". Investigative radiology, vol. 9, pp. 479-486, 1974.

[Rid92] H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments". Proceedings of the SPIE vol. 1666: Human vision, visual processing, and digital display III, pp. 16-26, 1992.

[RV93] M. de Ridder and J.P.A. Verbunt, "De invloed van ruis en onscherpte op beeldkwaliteit", IPO report no. 956, 1993 (in Dutch).

[RKY95] H. Roehrig, E. Krupinski and T. Yu, "Physical and psycho-physical evaluation of digital systems for mammography", Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 124-134, 1995.

[RC73] J.G Rogers and W.L. Carel, "Development of design criteria for sensor displays". Annual Report of Hughes Aircraft Comp., HAC ref. no. C6619, JAN-AIR, 1973.

[RJP87] D.C. Rogers, R.E. Johnston and S.M. Pizer, "Effect of ambient light on electronically displayed medical images as measured by luminance-discrimination thresholds". J. Opt. Soc. Am. A, vol. 4, no. 5, pp. 976-983, 1987.

[RMH95] J.P. Rolland, K. Muller and C.S. Helvig, "Visual search in medical images: a new methodology to quantify saliency". Proceedings of SPIE Medical Imaging 1995, SPIE vol. 2436: Image perception. pp. 40-48, 1995.

[RB92] J.P. Rolland and H.H. Barrett, "Effect of random background inhomogeneity on observer detection performance". J. Opt. Soc. Am. A, vol. 9, pp. 649-658, 1992.

[Ros73] A. Rose, *Vision: human and electronic*. Plenum Press, New York - London, 1973.

[Rou92] J.A.J. Roufs, "Perceptual image quality: concept and measurement". Philips Journal of Research, vol. 47, pp. 35-62, 1992.

[RB91] J.A.J. Roufs and M.C. Boschman, "Visual comfort and performance". In: J.R. Cronley-Dillon (gen. ed.), *Vision and visual dysfunction*, vol. 15: J.A.J. Roufs (ed.), Spatial vision. CRC Press Inc: Boston, 1991.

[RKT94] J.A.J. Roufs, V.J.F. Koselka and A.A.A.M. van Tongeren, "Global brightness contrast and the effect on perceptual image quality". Proceedings of the SPIE vol. 2179: Human vision, visual processing, and digital display V, pp. 80-89, 1994.

[SCH89] J.A. Saghri, P.S. Cheatham and A. Habibi, "Image quality measure based on a human visual system model". Optical Engineering, vol. 28, no. 7, pp. 813-818, 1989.

[Sch73] A.D. Schnitzler, "Analysis of noise-required contrast and modulation". In: L.M. Biberman (ed.), *Perception of displayed information.* Plenum Press, New York - London, 1973.

[Sch93] G. Schouten, *Luminance-brightness mapping: the missing decades.* Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1993.

[SRN+78] G.W. Seeley, H. Roehrig, S. Nudelman, M.M. Frost, and M.P. Capp, "Psychophysical evaluation of video systems compared to film". J. Appl. Photogr. Eng., vol. 4, no. 4, pp. 160-165, 1978.

[SYD87] M.I. Sezan, K.L. Yip and S.J. Daly, "An investigation of the effects of uniform perceptual quantization in the context of digital radiography". Proceedings of SPIE Medical Imaging 1987, SPIE vol. 767, 1987.

[Smi62] K. U. Smith, *Delayed sensory feedback and behavior.* W.B. Saunders Company, Philadelphia - London, 1962.

[SB80] W.M. Smith and K.F. Bowen, "The effects of delayed and displaced visual feedback on motor control". Journal of motor behavior, vol. 12, no. 2, pp. 91-101, 1980.

[SMLG75] S.J. Starr, C.E. Metz, L.B. Lusted and D.J. Goodenough, "Visual detection and localization of radiographic images". Diagnostic Radiology, vol. 116, pp. 533-538, 1975.

[STW+94] L.B. Stelmach, W.J. Tam, L. Wang, D. Lauzon and P. Gray, "Visual temporal masking at scene cuts". Perception vol 23, Supplement: Proceedings of the European Conference on Visual Processing '94, p. 10, 1994.

[SJ72] C.F. Stromeyer III and B. Julesz, "Spatial-frequency masking in vision: critical bands and spread of masking". J. Opt. Soc. Am., vol. 62, no. 10, pp. 1221-1232, 1972.

[Swe93] R. Swensson, "Measuring detection and localization performance". In: H.H. Barrett and A.F. Gmitro (Eds.), *Information processing in medical imaging.* 13th Conference IPMI, Flagstaff, Arizona. Springer: Berlin, 1993.

[SPW+79] J.A. Swets, R.M. Pickett, S.F. Whitehead, D.J. Getty, J.A. Schnur, J.B. Swets and B.A. Freedman, "Assessment of diagnostic technologies". Science, vol. 205, no. 4407, pp. 753-759, 1979.

[Val83] J.P. De Valk, *On the evaluation of medical images.* Ph.D. Thesis, Catholic University of Nijmegen, 1983.

[VE84] J.P.J. de Valk and E.G.J. Eijkman, "Analysis of eye fixations during the diagnostic interpretation of chest radiographs". Med. & Biol. Eng. & Comp., pp. 353-360, 1984.

[WB85] R.F. Wagner And D.G. Brown, "Unified SNR analysis of medical imaging systems". Phys. Med. Biol., vol. 30, no. 6, pp. 489-518, 1985.

[Wes91] J.H.D.M. Westerink, *Perceived sharpness in static and moving images.* Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1991.

[WT95] J.H.D.M. Westerink and C. Teunissen, "Perceived sharpness in complex moving images". Displays, vol. 16, no. 2, pp. 89-97, 1995.

[Wil91] H.R. Wilson, "Psychophysical models of spatial vision and hyperacuity". In: J.R. Cronley-Dillon (gen. ed.), *Vision and visual dysfunction,* vol. 10: D. Regan (ed.), Spatial vision. CRC Press Inc.: Boston, 1991.

[ZB80] E. van der Zee and R. Boesten, "The influence of luminance and size on the image quality of complex scenes". IPO annual progress report no. 15, pp. 69-75, 1980.