Generalizing the Evaluation of Medical Image Processing Tools by Use of Gabor Patterns

D. Apelt[1]; H. Strasburger[2]; R. Rascher-Friesenhausen[3]; J. Klein[4]; B. Preim[5]
[1]MeVis BreastCare Solutions GmbH & Co. KG, Bremen, Germany;
[2]Department of Medical Psychology and Medical Sociology, University of Göttingen, Göttingen, Germany;
[3]University of Applied Sciences, Bremerhaven, Germany;
[4]Fraunhofer MEVIS, Bremen, Germany;
[5]Otto-von-Guericke University, Magdeburg, Germany

Keywords
Perception, mammography, research design

Summary

Objective: Tools for medical image processing are usually evaluated by observers with radiological experience and with complex tasks. For easing evaluation of filtering and enhancement tools, the observer's task can be generalized.

Methods: By describing aspects of the MCS method (**M**ammographic **C**ontrast **S**ensitivity) we illustrate issues of selecting a metric for assessing visual performance, the observer's task and the image material to be used, aiming at a generalization of the design of studies for the evaluation of medical image processing tools. Concerning the metric, we distinguish acuity from contrast sensitivity. With respect to the observer's task, we distinguish tasks of discrimination from those at a higher level of recognition. Finally, we show the advantage of using medical images for evaluating image processing tools by comparing the results for measurements on homogeneous background and mammographic images.

Results: The perceptual level of the observer's task and the complexity of the used image material influences the outcome of observer studies, particularly also from crowding effects. The design of a study should minimize the impact of the observer's experience on the outcome. This can be achieved by using non-anatomical, standardized perceptual targets like Gabor patterns, used in the context of medical images.

Conclusions: Understanding the concepts of perception helps designing observer studies that are as complex as required, but at the same time as simple and general as possible. Performing an observer study may be simplified by a study design which does not require radiological experience of the observers, if the study aims at the evaluation of tools that shall support basic perception tasks, such as e.g. contrast enhancement.

Correspondence to:
Dörte Apelt
MeVis BreastCare Solutions GmbH & Co. KG
Universitätsallee 29
28359 Bremen
Germany
E-mail: doerte.apelt@gmx.de

Methods Inf Med 2009; 48: 331–335
doi: 10.3414/ME9231
prepublished: June 5, 2009

Introduction

The value of medical imaging and image processing methods is not only judged by the quality of the images, but ultimately by the radiologist's success [1]. Medical images are characterized by a large variability of structures, and interpreting those images is a challenge. Notwithstanding the diagnostician's expertise, lesions are missed. Reasons for not detecting a lesion can be grouped into three categories: an inappropriate search strategy so that the lesion is not looked at, not spending enough time for looking at the lesion, or, finally, misinterpreting the features of the lesion [2].

One aim of medical image processing is to help the radiologist look more carefully. The functionalities of available tools range from optimization of the workflow over contrast modifications, magnifications of image parts, and image enhancement strategies up to CAD (Computer Aided Detection) and CADx (Computer Aided Diagnosis) methods.

The subjective notion of "image quality" is influenced by image-dependent factors, related to the visual conspicuity of clinically relevant features, and on image-independent factors that are primarily semantic in nature and are related to what the observer knows about the presented visual information [1]. Tools for medical image processing and presentation can address both image-dependent and image-independent factors, e.g. by extraction, classification and marking of lesion structures with CAD methods (image-dependent), or by guiding the observer's visual-spatial attention, e.g. with display shutters (image-independent).

Studies evaluating the effect of image processing tools on the observer's performance are usually designed for observers with radiological experience [3–7]. Normally, the observer's tasks are of high complexity as, e.g., the detection or discrimination of a (simulated) lesion in a mammographic image [3, 8]. The effort for the development of the observer's task, for selecting the image material and for selecting the observers is high. Here we show that evaluation studies for e.g. contrast enhancement tools might be supported by methods adapted from basic perceptual science, especially with regard to a generalization of the observer's task.

Objectives

Three perceptual issues are discussed: the choice of a metric for describing the observer's visual performance, the observer's perceptual task, and the image material used. The observer's task and the metric for assess-

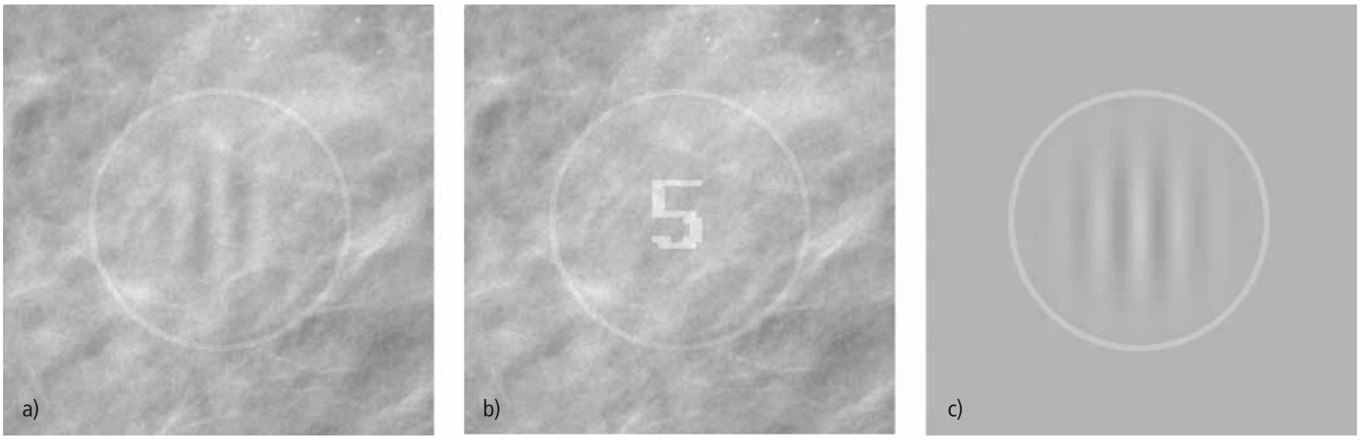

Fig. 1 Examples for the presented test items in the investigation of the MCS method, superimposed onto a mammogram and a homogeneous image, respectively: a) and c) Gabor pattern; b) digit. The surrounding circle marks the area to be fixated by the observer [9].

ment of visual performance are coherent and strongly affect the effort of an observer study. Discussing the selection of image material is essential for investigating evaluation methods in medical image processing.

Exemplified by the MCS method (Mammographic Contrast Sensitivity) [9–11], an approach for a more general study design is illustrated. The MCS method was developed for investigating the effect of tools on a special performance measure, the individual contrast sensitivity on mammographic backgrounds. Our motivation for the development of the MCS method has been to make the study results independent of the observer's radiological expertise.

Methods

The MCS Method as an Example of a Perception-centered Approach

The MCS method uses Gabor patterns superimposed onto a mammographic image background for the determination of the individual contrast sensitivity [9]. Gabor patterns are sinewave-modulated "stripes" in a Gaussian soft aperture (▶Fig. 1c). The observers are given an orientation discrimination task in which they decide, e.g., whether the "stripes" are horizontal or vertical. Contrast thresholds are measured by a psychophysical staircase procedure [12] at six spatial frequencies up to 16 cycles per degree (cpd; number of "stripes" per degree visual angle). The use of Gabor patterns, overlaid onto a mammogram, is motivated by the fact that spatial frequency content does influence the radiologist's perception and interpretation of images [2, 13–15], and by that Gabor patterns are today's standard stimulus for assessing human contrast sensitivity.

The measurements can be performed under different viewing conditions or with a number of image presentation or processing tools. The results from these manipulations show whether the particular conditions improve the observer's contrast resolution performance. ▶Figure 1a shows a Gabor pattern on a mammogram, and ▶Figure 2 shows the plot of the contrast sensitivity function (CSF) for orientation discrimination, measured with a set of Gabor patterns of differing spatial frequency. For the MCS method the contrast sensitivity is defined as usual with Gabor stimuli as the inverse of the Michelson contrast $(L_1 - L_2)/(L_1 + L_2)$, where L_1 and L_2 are the minimum and maximum luminance in the Gabor pattern, respectively.

These results were obtained in an observer study, conducted to verify the applicability of the MCS method in the mammographical context [9]. Eight observers took part in the psychophysical study. On homogeneous images as well as on four mammograms having different tissue characteristics, Gabor patterns and digits, respectively, were presented on an area of about 2.5 cycles per degree and for a time of 720 ms (Fig. 1). The location of the Gabor patterns and digits was marked by a fixation circle. The observer's task was to focus on the given area, to determine the orientation of the Gabor patterns, and to identify the digits, respectively. Contrast levels and spatial frequencies were varied. The results of the study are used in the following sections.

Metrics for Visual Resolution Performance

Visual performance is standardly assessed, amongst other measures, by visual acuity and contrast sensitivity. Acuity specifies the minimum separable distance between two contours at maximum contrast. Acuity, therefore, provides a suitable description of visual performance for investigations where fine detail at maximal contrast plays an important role. Medical images are, however, often characterized by the presence of both coarse and fine detail and structure of low contrast.

To account for the latter, the use of contrast sensitivity as a metric is advisable. Contrast sensitivity is the inverse of the contrast threshold, where threshold is the minimum contrast required for detecting or identifying a pattern in front of a given background. Usually, the contrast sensitivity is determined as a function of spatial frequency, resulting in the contrast sensitivity function (CSF; Fig. 2) [15]. The CSF is a more comprehensive description than a single contrast sensitivity value, since the perception of coarse structure (low spatial frequency) cannot be predicted from the perception of finegrain structure (high spatial frequency). In general, the perception of low and high spatial frequency requires more contrast than at

 © Schattauer 2009

the mid range [8]. Principal component analysis has shown that most of the CSF's variation is captured by three factors, or bandpass channels, peaking at low, medium and high spatial frequency [16].

The Observer's Task

Human pattern recognition is a nonlinear process [17, 18], and the results of contrast threshold measurements, performed e.g. with a detection task and a homogeneous image background, might not allow predictions on the sensitivity for recognizing low-contrast patterns in an applied situation as in the diagnostics of mammograms. Indeed, previous psychophysical studies have shown that human detection performance does not predict pattern recognition performance [19]. So, for tasks in visual perception, a processing hierarchy is assumed: The perception of a structure can occur at different levels, such as detection, discrimination, identification and recognition, in an ascending order of hierarchical relation. It is not always possible to predict results at one hierarchy level from those at another level [19, 20]. If an observer detects a structure, there is no prediction of whether the observer will be able to discriminate the structure from another one, or whether he or she is capable of recognizing the structure.

This difference can be illustrated by comparing the results of the MCS method with those of a modified MCS method, i.e. where the task of discriminating Gabor patterns is replaced by one of identifying a pattern, like a digit in our application (▶Fig. 1b). ▶Figure 3 shows the results of the pooled data for the eight observers of the study mentioned above. The CSF for the more difficult task – the identification of a digit – is lower than the CSF for the easier task of orientation discrimination. This behavior is found for measurements with homogeneous background as well as for those on mammographic background images.

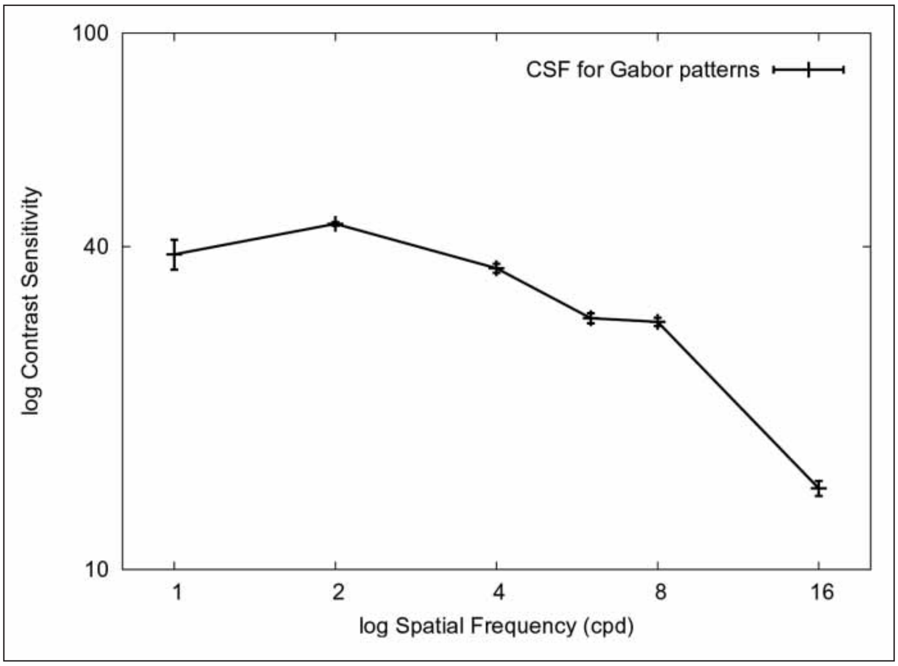

Fig. 2 Plot of the contrast sensitivity function (CSF) for the orientation discrimination with a set of Gabor patterns of different spatial frequencies. The presented CSF includes the pooled CSF data of the eight observers (modified from [9]).

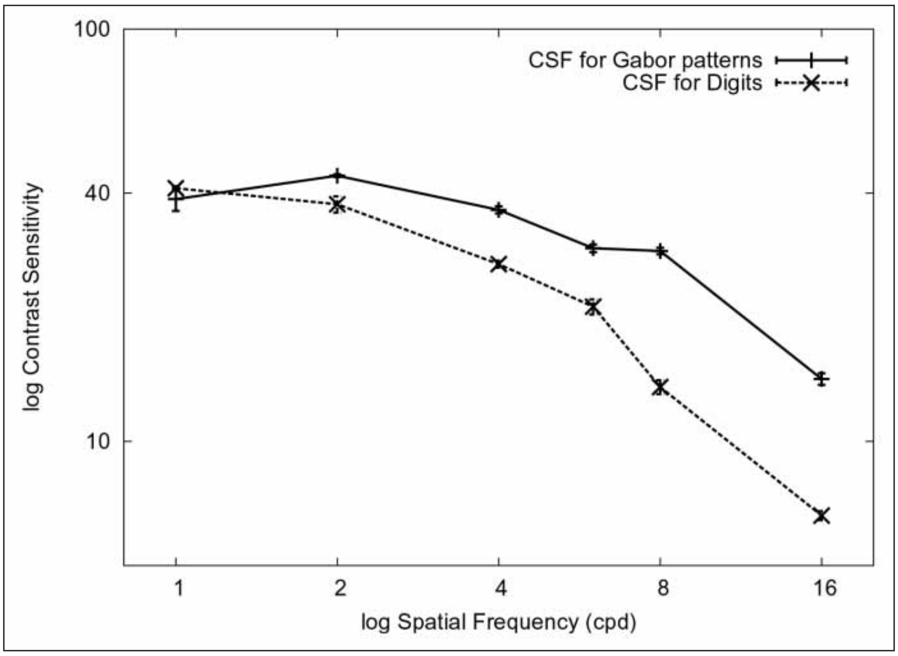

Fig. 3 Comparison of the CSF for a discrimination task with Gabor patterns (continuous line) and an identification task with digits (dashed line) for eight observers (modified from [9])

Image Material for Perceptual Investigations

The identification of an object within a medical image is influenced both by its image features and by the features of the surrounding anatomical structures. In basic perceptual research the impairment of pattern recognition by the presence of neighboring structure is known as the crowding effect [21–23]. Samei et al. demonstrated this effect for the detection of subtle lesions in X-ray images of the thorax [24]. Investigations on observer's performance for tasks on medical images are thus expected to have higher validity when

© Schattauer 2009

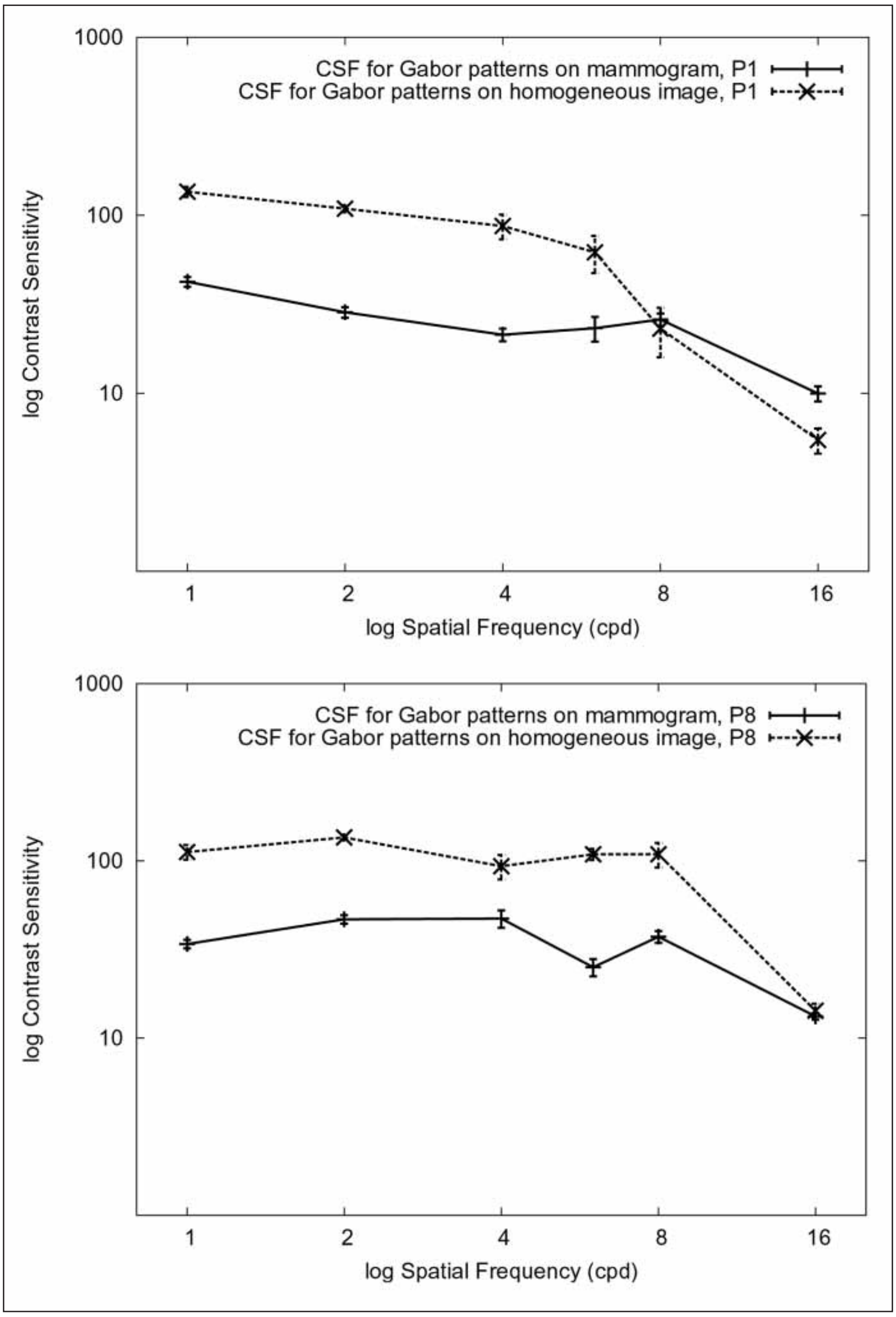

Fig. 4 Comparison of the CSF of two observers (P1 and P8) for Gabor patterns on homogeneous background (dashed line) and a mammographic image (continuous line) (modified from [9])

using objects on real medical instead of homogeneous images.

For investigation of the behavior of the CSF in an anatomical and a non-anatomical image context, the MCS method was performed with a mammographic and a homogeneous image background (▶Figs. 1a and 1c). For all observers the contrast sensitivity was significantly higher on the homogeneous than on the tissue background. An example of this for the observers P1 and P8 is shown in ▶Figure 4. Not only the overall level but also the shape of the CSF varied between the homogeneous and the tissue image context. These variations were different between the observers. The precise shape of the CSF on the tissue background thus seems not easily predictable from the CSF on a homogeneous background.

Results and Discussion

To sum up, for investigations on the observer's performance, basic perceptual science shows the role of the selection of the task level and of the background. The use of real medical images (over homogeneous or simple phantom images) can be expected to improve validity, since the anatomical and pathological variations in an image influence the observer's performance, e.g. due to masking or crowding effects. There is, furthermore, a range of object sizes for which the perceptual performance is better than for smaller and larger objects [8].

For the application of the MCS method this suggests to use mammographic images to meet the medical context, but to make the investigation more general and independent from the observer's expertise by giving the observer a task which is perceptual rather than cognitive in nature to find a lesion without knowing its structural details. The task of orientation discrimination for Gabor patterns with a varying spatial frequency, which is standard in vision research, meets this requirement. The higher-level task of identifying a letter at low contrast can furthermore be expected to give results closer to the recognition of low-contrast anatomical aberrations, if this is required by the tool to be evaluated.

When non-anatomical perceptual targets such as Gabor patterns or characters are chosen that do not require radiological experience, the results can be expected to reflect general perceptual characteristics only. However, if a medical image processing tool aims at the improvement of general perceptual characteristics, as the perception of contrast, the use of non-anatomical targets may reduce the effort of tool evaluation studies with regard to the selection of observers and the preparation of image material. Moreover, the reproducibility and reliability of the measurement results may be improved, since different radiological expertise between the observers has not to be considered.

Conclusions

Since technology allows presenting better images, perceptual research in medical imaging

© Schattauer 2009

may gain importance. Results from image perception may guide basic research on the efficacy of image processing tools [25] and on the evaluation of their effects on selected perceptual performance aspects of the observer. Methods adapted from basic perception research can complement methods of tool evaluation that use pure mammographic image material and tasks.

The investigations on the MCS method indicate how to generalize tool evaluation by simplifying the target item and observers' task. Although the CSF for more complex targets like characters may provide results closer to radiological perception tasks, studies using Gabor patterns can be used for general statements on the effect of image processing tools, i.e. whether the application of a tool leads to an improved contrast perception for a set of spatial frequencies.

A crucial open issue is how to simplify targets and observers' tasks without compromising the medical relevance of the outcome. Further investigations are necessary for an explicit comparison of the results of studies using generalized and anatomical targets. If for example an improvement of contrast perception is achieved for several types of targets, we can expect that the generalized targets are sufficient for the tool evaluation.

In addition, further investigations should be made on the correlations between different target items, such as Gabor patterns, digits and simulated simplified lesions, and the mammographic image background. In another investigational part of the study, described in [9], the results of the MCS method with Gabor patterns showed a wide independence of the mammographic image. An interesting question is whether this independence also exists for more difficult target items.

When designing and performing observer studies, results from human pattern recognition research point to the importance of which metric to choose for assessing the visual performance, to the level of the observers' task and to the used image material.

References

1. Manning DJ, Gale A, Krupinski EA. Perception research in medical imaging. Br J Radiol 2005; 78: 683–685.
2. Krupinski EA. The future of image perception in radiology: Synergy between humans and computers. Acad Radiol 2003; 10: 1–3.
3. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. Med Phys 2001; 28 (4): 419–437.
4. Awai K, Murao K, Ozawa A, Komi M, Hayakawa H, Hori S, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. Radiology 2004; 230: 347–352.
5. Zimmerman JB, Pizer SM, Staab EV, Perry JR, McCartney W, Brenton BC. An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. IEEE Trans Med Imaging 1988; 7 (4): 304–312.
6. Graf B, Simon U, Eickmeyer F, Fiedler V. IK versus 2K monitor: a clinical alternative free-response Receiver operating characteristic study of observer performance using pulmonary nodules. AJR Am J Roentgenol 2000; 174: 1067–1074.
7. Cederberg RA, Frederiksen NL, Benson BW, Shulman JD. Influence of the digital image display monitor on observer performance. Dentomaxillofac Radiol 1999; 28 (4): 203–207.
8. Burgess AE, Jacobson FL, Judy PF. Detection of lesions in mammographic structure. Proceedings of SPIE 1999; 3663: 304–315.
9. Apelt D, Peitgen HO. Contrast sensitivity in mammographic softcopy reading – determination with psychophysical procedures. Digital Mammography, 9th International Workshop, IWDM 2008: Tucson, AZ, USA: Springer LNCS 5116: 756–763.
10. Apelt D, Strasburger H, Rascher-Friesenhausen R, Klein J, Preim B, Peitgen HO. Contrast sensitivity in mammographic softcopy reading. Proceedings of SPIE 2009; 726318.1–726318.11.
11. Apelt D, Rascher-Friesenhausen R, Klein J, Strasburger H, Preim B, Peitgen HO. Impact of luminance distribution in the visual field on foveal contrast sensitivity in the context of mammographic softcopy reading. Proceedings of SPIE. 2009; 72630B.1–72630B.9.
12. Treutwein B. Adaptive psychophysical procedures. Minireview. Vision Res 1995; 35: 2503–2522.
13. Mello-Thoms C. Perception of breast cancer: eye-position analysis of mammogram interpretation. Acad Radiol 2003; 10: 4–12.
14. Mello-Thoms C, Chapman B. A preliminary report on the role of spatial frequency analysis in the perception of breast cancers missed at mammography screening. Acad Radiol. 2004; 11: 894–908.
15. Campbell FW, Robson JG. Application of Fourier analysis to the visibility of gratings. J Physiol 1968; 197: 551–566.
16. Simpson WA, McFadden SM. Spatial frequency channels derived from individual differences. Vision Res 2005; 45: 2723–2727.
17. Zetzsche C, Barth E. Fundamental limits of linear filters in the visual processing of two-dimensional signals. Vision Res 1990; 30: 1111–1117.
18. Barth E, Zetzsche C, Rentschler I. Intrinsic two-dimensional features as textons. J Opt Soc Am A 1998; 15: 1723–1732.
19. Strasburger H, Rentschler I. Contrast-dependent dissociation of visual recognition and detection fields. Eur J Neurosci 1996; 8: 1787–1791.
20. Strasburger H. Indirect vision [Indirektes Sehen]. (in German). Göttingen: Hogrefe; 2003.
21. Levi DM. Crowding – an essential bottleneck for object recognition: a mini-review. Vision Res 2008; 48 (5): 635–654.
22. Strasburger H, Harvey LO Jr, Rentschler I. Contrast thresholds for identification of numeric characters in direct and excentric view. Percept Psychophys 1991; 49: 495–508.
23. Strasburger H. Unfocussed spatial attention underlies the crowding effect in indirect form vision. J Vis 2005; 5 (11): 1024–1037.
24. Samei E, Flynn MJ, Peterson E, Eyler WR. Subtle lung nodules. Influence of local anatomic variations on detection. Radiology 2003; 228: 76–84.
25. Krupinski EA. The importance of perception research in medical imaging. Radiat Med 2000; 18 (6): 329–334.

© Schattauer 2009