# Contrast sensitivity and discrimination in pictorial images

Triantaphillidou, S., Jarvis, J. and Gupta, G.

This is a pre-publication version of a paper published in: SPIE Proceedings 9016, Image Quality and System Performance XI, SPIE. The final definitive version is available from the publisher at:

https://dx.doi.org/10.1117/12.2040007

# Spatial contrast sensitivity and discrimination in pictorial images

Sophie Triantaphillidou*, John Jarvis, Gaurav Gupta
Imaging Technology Research Group, School of Media, Arts and Design,
University of Westminster, Watford Road, HA1 3TP, Harrow, UK

## ABSTRACT

This paper describes continuing research concerned with the measurement and modeling of human spatial contrast sensitivity and discrimination functions, using complex pictorial stimuli. The relevance of such functions in image quality modeling is also reviewed. Previously[1,2] we presented the choice of suitable contrast metrics, apparatus and laboratory set-up, the stimuli acquisition and manipulation, the methodology employed in the subjective tests and initial findings. Here we present our experimental paradigm, the measurement and modeling of the following visual response functions: i) *Isolated Contrast Sensitivity Function* (iCSF); *Contextual Contrast Sensitivity Function* (cCSF); *Isolated Visual Perception Function* (iVPF); C*ontextual Visual Perception Function* (cVPF). Results indicate that the measured cCSFs are lower in magnitude than the iCSFs and flatter in profile. Measured iVPFs, cVPFs and cCSFs are shown to have similar profiles. Barten's *contrast detection* model[3] was shown to successfully predict iCSF. For a given frequency band, the reduction, or masking of cCSF compared with iCSF sensitivity is predicted from the *linear amplification model* (LAM)[4]. We also show that our extension of Barten's *contrast discrimination* model[1,5] is capable of describing iVPFs and cVPFs. We finally reflect on the possible implications of the measured and modeled profiles of cCSF and cVPF to image quality modeling.

**Keywords:** Contrast sensitivity, contextual contrast sensitivity, contrast discrimination, contextual contrast discrimination, Barten detection model, linear amplification model (LAM), image quality, image fidelity, CSF implementation.

## 1. INTRODUCTION

The paper describes the most recent developments of our research, concerned with the determination of human contrast sensitivity and contrast discrimination functions from complex pictorial stimuli. We maintain that such functions are relevant to image fidelity and image quality modeling, and that findings from our research may provide some answers on the relevance and application of the Contrast Sensitivity Function (CSF) in fidelity or/and quality models. We measure two sensitivity functions, each at two different conditions[1,2]: the *Isolated Contrast Sensitivity Function* (iCSF) and the *Isolated Visual Perception Function* (iVPF) are contrast detection (threshold contrast sensitivity) and discrimination (suprathreshold contrast sensitivity) functions respectively, measured from bandpass images *in isolation* (i.e. displayed in uniform luminance backgrounds). The *Contextual Contrast Sensitivity Function* (cCSFs) and the C*ontextual Visual Perception Function* (cVPFs) are again contrast detection and discrimination functions respectively, but measured from bandbass images *in context* (i.e. displayed as embedded in the pictorial images to which they belong).

Previously[1,2] we presented the choice of suitable contrast metrics, apparatus and laboratory set-up, the stimuli acquisition and manipulation, and the methodology employed in the subjective tests. Here, a short review of the use and implementation of CSFs to different genres of image quality and fidelity models is given in Section 2. Section 3 presents our revised experimental paradigm, the range of stimulus conditions and defines Barten's detection[3] and discrimination[5] models, which are used in our study. We have extended the latter model, to address spatial frequency[1]. In Section 4 measured and modeled results from one pictorial image, at two different contrast levels, are presented and discussed, along with initial thoughts on possible implications of the findings in image quality modeling. Conclusions are drawn in Section 5. The appendix explains our derivation of the theoretical relationship between contextual detection (cCSF) and isolated detection (iCSF) through application of the linear amplification model (LAM)[4].

*triants@westminster.ac.uk; phone +44 (0)20 7911 5000; fax +44 (0)20 7911 5943

## 2. THE RELEVANCE OF CONTRAST DETECTION AND DISCRIMINATION MODELS TO THE MODELING OF IMAGE QUALITY AND FIDELITY

One approach to image quality modeling has its origins in traditional imaging science, with its roots in chemical imaging, where quality (defined as *the subjective –visual- impression the image conveys*) is assumed to have a functional relationship with measured attributes in imaging systems (e.g. measures of sharpness, noise, etc.) Image quality metrics/models originating from this school of thought combine one or more imaging performance measures (e.g. MTF, PSF, NPS, etc.) with a model of the Human Visual System (HVS), typically the Contrast Sensitivity Function (CSF), to predict the visual quality, as judged by human observers, i.e. as measured using psychophysics. Examples of these types of metrics are: SQF[6], MTFA[7], SQRI[8], PIC[9], and more recently the EPIC[10] metric, with the SQRI (a metric determined by integrating the system's spatial characteristics and the CSF as a weighting function) being the most widely accepted paradigm. This is rather an "engineering approach" to image quality, dating back to Schade et al.[11], who first extended optical system descriptions, such as the MTF, to the visual process. In this modeling approach, the output image is considered as function of the input signal (input intensities), the physical characteristics/limitations of the imaging system components (capture, image processing, print/display) as well as the human visual system; and all three are separable. An advantage of this type of approach is that, the performance of the imaging chain, and thus the quality associated with its imagery, can be cascaded from the measured performance of the individual imaging component systems/processes (and vice versa). Also, the model algorithms are simple to implement. A considerable disadvantage however is that, the performance of the imaging components (including that of the HVS) is measured using simple test patterns, such as gratings, edges, and noise patterns. Thus relevant image quality metrics fail to address scene dependency, i.e. the variation of the human responses, as well as the variation in the performance of imaging algorithms with different types of complex signal, that is different types of pictorial imagery [12-14]. Further, there have been great debates over whether the non-linear contrast transduction by the HVS can be described by the CSF[15] and cascaded along with measures linked to linear system theory[16,17]. The engineering type of metrics are considered to only be suitable for the valuation of image quality optical systems, the visibility of uniform noise or of uniformly distributed artifacts[18].

During the late 1980's, another outlook to quality modeling was already being in development. This approach was born with digital imaging and addressed image fidelity (defined *as the ability to quantify detection and visual differences between images)*. The first models were designed to predict visual differences between two images (typically but not exclusively an original and a compressed version). The 1D CSF was considered too simple a model to describe alone all complex functions of the HVS in viewing complex scenes. The non-linearities of various digital image processes were undoubtedly not compatible with linear system theory. Along with the use of 2D CSFs, other processes accounting for visual masking, contrast gain, local contrast, oriented responses and probability of detection were incorporated in the image fidelity models of the time. These models use images, rather than parameters of the imaging system in fidelity predictions. They tackle directly problems related to visual masking encouraged when viewing pictorial imagery, as well as the scene dependency of imaging algorithms. Typically, the images are frequency decomposed to different visual channels and differences between the weighted channels of the original and the distorted image versions are pooled. Their development was facilitated by the fact that images were now made out of numbers and thus a "computational approach" that included many complicated visual processes and system simulations was more accessible. One of the most eminent of these fidelity models was Daly's VDP[16]. In this a 2D bandpass CSF is used, prior to frequency decomposition, as a weighting function having many parameters, including radial spatial frequency, luminance, image size, and eccentricity and viewing distance. In another, Lubin's mechanistic model for detection and discrimination[19] developed to quantify the visibility of display information, the CSF is used as a base sensitivity normalization factor after channel decomposition. Some image difference models use weightings of the spatial frequency channels derived empirically, instead of relying on a threshold sensitivity model such as the CSF, or no CSF at all[20].

Throughout the advances in quality and fidelity modeling, debates have been published on whether threshold contrast sensitivity is relevant to image quality, since the latter deals with suprathreshold contrasts. But conflicting views and evidence on both its relevance and implementation have continually emerged. Notably, Silverstein and Farrell in 1996[21] found weak correlations between image fidelity and image quality and declared that suprathreshold judgments are unrelated to contrast threshold judgments. They stressed that we should not expect human vision models based on contrast detection to predict judgments of image quality, but that it is possible to develop computational measures of image fidelity using them. But, they recognized the role of contrast detection in image quality, since the detection of artifacts (for example those introduced by compression) impacts its judgment.

Other than in the traditional image quality models and some fidelity models designed for compression, the CSF has been reported successful when incorporated in models dealing with visibility of halftone patterns (eg. Zhang and Wandell)[22], in image rendering and image appearance models that focus both on the prediction of image quality scales, as well as suprathreshold visual differences (e.g Fairchild and Johnson)[23,24]. Opposite views have come from many too. Ahumada and his colleagues[20,25], after investigating various computational models concluded that, CSF weighting does not necessarily improve image quality metrics designed to predict the detectability of objects in complex backgrounds. Rogowitz et al.[26] questioned the suitability of simple detection models applied to suprathreshold scaling methods for evaluating the perceived quality of images with multiple suprathreshold artifacts.

Some recent successful approaches to image fidelity and quality do not use low-level visual descriptors at all. Instead they operate on various hypotheses on what the HVS attempts to accomplish when presented with pictorial imagery. SSIM[14,27] calculates structural similarity between original and distorted image versions and operates on the assumption that the HVS has evolved to extract structural information from natural scenes. The ICF model[28], an information theoretic approach to fidelity, is based on the assumption that human vision has evolved according to the statistical properties of the natural environment, and thus a distorted image can be quantified according to the amount of information it provides about the original. An extension of ICF, the VIF[29], has accounted for some properties of vision, but is missing low-level properties in the name of analytical and computational simplicity and independence from viewing conditions. It collects all sources of HVS uncertainty into one additive noise component, thus expressing the CSF in terms of equivalent internal noise. Because these models ignore the HVS response under different viewing conditions (e.g. viewing distance) and do not account for any output system characteristics (e.g. display MTF and luminance non-linearity) their results cannot be calibrated for different viewing conditions, or systems. In parallel to the development of such models, current successful models, such as the VSNR[30] amongst numerous others, still incorporate a CSF.

Haun and Peli[31] recently argued that in estimating the visual quality of an image, contrast thresholds are of principal importance, whilst perceived suprathreshold magnitudes are relatively less important; and that the specific sensitivity functions commonly used to estimate perceived contrast and quality may be misapplied, or inappropriate. So how should contrast detection models be applied? The most common approaches have been to weight the signal by applying the CSF as a spatial filter, either in the spatial or spatial frequency domain; or after channel decomposition, to weight the channel sensitivity so that the sum approaches the target CSF. Peli[32] tested the validity of the threshold CSF derived from Gabor patches as a weighting function in predicting sharpness differences and found that the best CSFs for image quality modeling are low pass, flat at low frequencies (i.e. not typical bandpass CSFs). In a late search of Haun and Peli[33] for contrast weightings relevant to the perception of contrast in suprathreshold complex imagery, they produced scene dependent perceived contrast weights, forming a bandpass spatial frequency function. They suggested that the perceptual impact of an image, and the way its contrast is interpreted by an observer, is dependent on the structure of the image as well as the type of observer.

Our view is that contrast perception models are relevant to image fidelity and image quality, especially when the latter decreases monotonically with the visual appearance of any type of "noise" (i.e. noise in its wider sense, relating to any information preventing visually optimum transmission of the signal). But questions remain to be answered. How do CSFs derived from simple stimuli relate to the viewing of complex pictorial imagery? Are contrast discrimination functions (VPFs) more relevant to suprathreshold image contrasts than the CSF? Are functions other than CSFs or VPFs more suitable weighting functions for application in image quality modeling? According to Haun et al.[15], the quest for a standard spatial observer, which can make both qualitative and quantitative judgments from images, is a very complex matter. The Modelfest experiment[34,35] has been an effort to establish a standard spatial observer relevant to image quality modeling, but has employed mainly non-pictorial stimuli. One of our aims is that, by extracting contrast sensitivities and discrimination functions directly from pictorial imagery, we may shed some light to some of these long contested matters.

# 2. METHODOLOGY

## 2.1 Experimental paradigm

As previously reported[2], we measure four visual response functions. All experiments involve the presentation of an unchanged *standard* stimulus and a modified *test* stimulus.

i) *Isolated Contrast Sensitivity* (iCSF), describes the ability of the visual system to detect any spatial signal in a given spatial frequency octave in isolation and is the closest equivalent to a conventional sinewave CSF. It is measured by presenting the standard, comprising a uniform field of mean luminance equal to the mean luminance of the image, against the test, comprising a variable increment of band contrast around the standard. The comparisons continue until the test image is perceived as just different from the standard. The value of the band contrast of the chosen test image is the *band contrast detection threshold.*

ii) *Contextual Contrast Sensitivity Function* (cCSF), describes the ability of the visual system to detect a spatial signal in a given octave contained within an image. The procedure for obtaining this measurement is the same as for the iCSF, but now both standard and test consist of an image containing all pictorial information outside the band of interest (i.e. the selected band is removed). As in the measurement of the iCSF, the test is shown with a variable increment of band contrast and the comparisons continue until the test is perceived as just different from the standard. Again, the value of the test band contrast is the *contrast detection threshold,* but this time defined for the contextual sensitivity conditions. A comparison between isolated and contextual band conditions indicates the extent that spatial information outside of the band of interest acts as a source of signal masking.

iii) *Isolated Visual Perception Function* (iVPF), describes visual sensitivity to changes in suprathreshold contrast of any spatial signal in a given spatial frequency octave in isolation. The procedure for obtaining this measurement is similar to that used for iCSF determination, but now both standard and test consist of a band in isolation, containing full pictorial information.

iv) C*ontextual Visual Perception Function* (cVPF), describes visual sensitivity to changes in suprathreshold contrast in an image. The procedure for obtaining this measurement is similar to that used for cCSF determination, but now both standard and test consist of an image containing full pictorial information in all bands. The contrast of the band of interest within the test image is progressively reduced until a difference between the test and the standard is just perceived. For a given band, the difference between the contrast of the standard and the contrast of the 'just perceived' test will produce the *contrast difference threshold*.

Figure 1 presents illustrative examples of the low contrast version of one test image from out image set (i.e. low contrast "Gallery"), for all experimental conditions i), ii), iii) and iv).

## 2.2 Choice and conditions of visual stimuli

Careful thought has been put into the collection of suitable visual stimuli for our experiments. The stimulus acquisition and calibration, as well as the characterization and setup of the imaging chain we used for the presentation of the stimuli to selected observers are detailed in reference 1. The stimulus conditions we aim to investigate, including variations in stimulus mean luminance, color information, noise and contrasts, along with the types of scenes we have selected for experimentation are listed in reference 2.

The stimulus size and viewing specifications for all conditions and stimuli are as follows: field size $15^{o}$ subtense; one mode of view: *global free-view scanning*; viewing distance 1.8m; spatial frequencies centered 0.1- 32 cycles/deg.

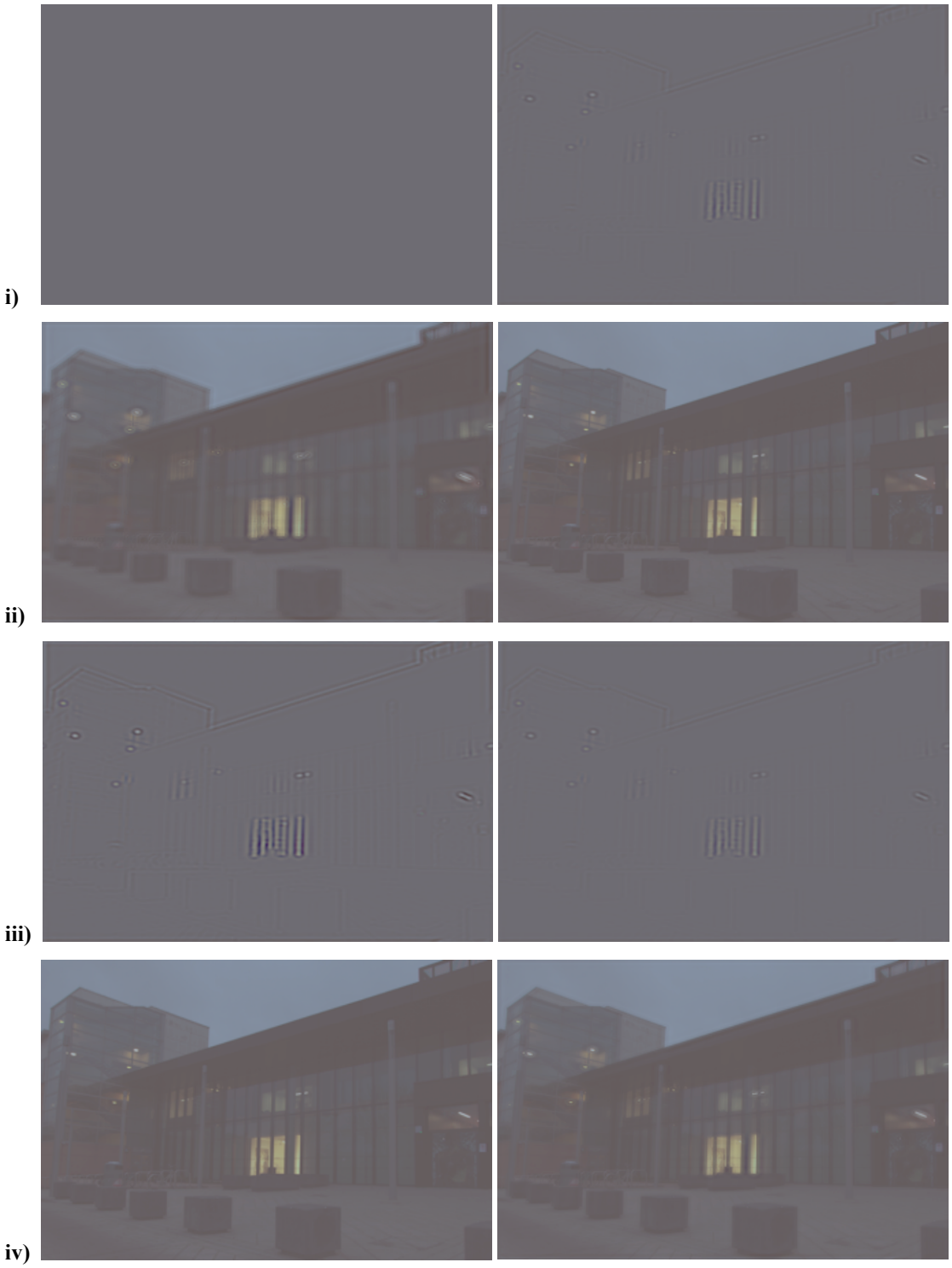


Figure 1: Standard images (left) and test images (right) presented for i) iCSF experiment, ii) cCSF experiment, iii) iVPF and iv) cVPF experiments, using the low contrast version of the “Gallery” image.

**2. 3 Modeling contrast detection and discrimination**

Barten has presented two models relevant to our studies in his classical thesis on contrast sensitivity: a model for contrast detection[3] and a model for contrast discrimination[5]. We extended the Barten model for contrast discrimination to provide a full description of the VPF for static images and demonstrated its capability for describing measured sensitivity data obtained from sinewave as well as complex image forms[1]. The extension involves the inclusion of spatial frequency as a variable.

In brief: the detection model expresses contrast detection, $c_t$, as a function of spatial frequency, $u$, in cycles per degree (c/deg), by:

$$c_t(u) = \left[ KO(u)H(u)A(u)\left(\Theta(u)\right)^{-0.5} \right]^{-1} \tag{1}$$

where, K is a constant, and the terms O,H,A and Θ are filter signal transfer characteristics associated with photon capture, lateral inhibition in the retina, and area summation respectively. The term Θ is a combination of both photon noise and internal neural noise. The numerical values of most parameter values in Equation 1 are already known[3,36,37] and the remainder, which are specific to particular stimulus conditions, can be readily calculated[38]. Our formulation for the VPF is given by:

$$\mathrm{VPF}(u) = \left[ \sqrt{\frac{c_t^2(u) + 0.04k^2c_s^2(u)}{1 + 0.004k(c_s(u)/c_t(u))} + c_s^2(u)} - c_s(u) \right]^{-1} \tag{2}$$

Equation 2 shows that if contrast at detection $c_t$ is known, the VPF can be immediately calculated for a given pedestal (or narrow band) contrast $c_s$. The constant k is the inverse of the Crozier coefficient[39], and represents the minimum signal/noise level required for detecting a spatial signal. In human vision, this is approximately equal to 3.0[11,40]. The inverse of $c_t$ gives the classic CSF. At a given frequency, the VPF represents the reciprocal of the difference between the contrast level of a suprathreshold sinewave signal ($c_s$) and a level ($c$), which can just be discriminated, i.e. VPF = $| c - c_s |$. Note that as pedestal contrast approaches zero, $\Delta c$ approaches $c_t$. In other words, both the CSF and VPF collapse into a single function.

As previously reported[2], initial measurements of both iCSF and cCSF were obtained from a chosen pictorial scene with a relatively low contrast spectrum. RMS band contrast was the metric chosen to calculate sensitivities. The results revealed an iCSF, which closely follows the profile and magnitude of a sinewave CSF, appropriate for the stimulus size and mean luminance employed. Measurements of cCSF were found to be both lower in magnitude and produce a flatter function in profile compared with the iCSF, showing that basic band detection is masked by the presence of the picture.

Measurements of iCSF and cCSF revealed that:

$$c_c^{\,2}(u) - c_i^{\,2}(u) = \text{constant } c_s^{\,2}(u) \tag{3}$$

where $c_c$ and $c_i$ denote the detection contrast ($c_t$) for contextual and isolated conditions.

It is well established that spatial sinewave detection is generally masked when the stimulus is embedded within other spatial information, whether this be in the form of white[41], or band filtered 2D noise[42,43], or 1D compound grating structures[44]. In most of these studies, the measured CSF reduces in magnitude, and its peak shifts to relatively high frequencies (10-20 c/deg). Again, previous work on sinewave detection suggests that the majority of this noise will originate from a frequency range of + and – 1 octave from the stimulus sinewave[2,44]. Therefore, a likely source for the detection suppression of a given octave band of the cCSF is noise generated from the flanking signal bands centered + and – 1 octaves away. With this assumption, the *linear amplification model* (LAM)[4], which relates the power spectral density of the masking noise to contrast detection, theoretically predicts Equation 3, as shown in the Appendix. Previous experimental studies of sinewave detection in the presence of noise have successfully employed the LAM to quantify the extent of the masking[42,45,46,47].

Equation 3 enables contextual detection to be calculated directly from the iCSF and the contrast spectrum of the image. Equations 1, 2 and 3 fully define our current model framework for describing detection and discrimination sensitivities in real scenes.

# 3. RESULTS AND DISCUSSION

Sensitivity measurements for the iCSF, cCSF, iVPF and cVPF, obtained for our first selected stimulus, the Gallery image (Figure 1), at two different contrast levels (see section 4.2, Figure 3), are presented and discussed below, along with model predictions. The measurements represent mean responses from minimum four and maximum eight selected observers[2].

## 3.1 iCSF

The data points in Figure 2 are measured values of iCSF for the two image contrasts. Error bars represent the standard mean error[48]. The curve represents Equation 1, with parameter values fixed for a 15-degree stimulus size and a mean luminance of 10 cd/sqm. Since the band contrast does not play a role in basic isolated detection, the measured iCSFs are the same for both the high and the low contrast versions of the stimulus. The fact that Equation 1 appears to model successfully the iCSF data is interesting, because the stimuli used for the iCSF experiment contain pictorial information only from a narrow band at a time (pictorial content not dissimilar to images of gratings) on an otherwise uniform background.

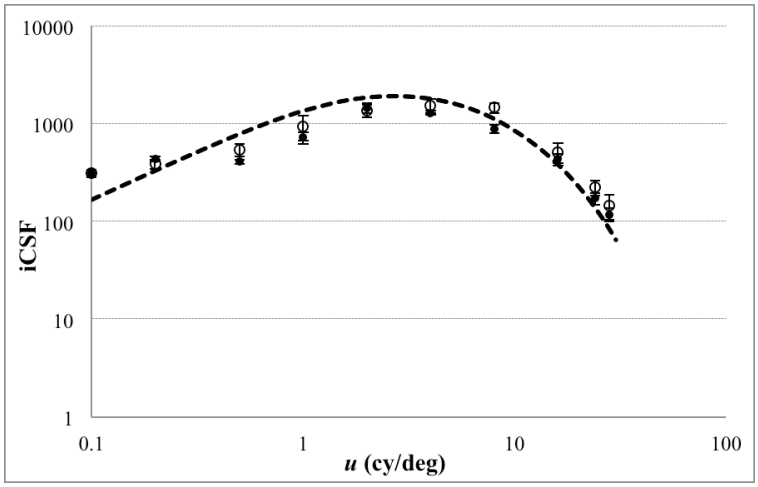


Figure 2. Measured (data points) and modeled iCSF. Black circles: low contrast image, white circles: high contrast image.

## 3.2 Image Contrast Spectra

Figure 3 shows image contrast spectra (expressed as $c_s^2$ ) for the low and high contrast versions of the Gallery image. For spatial frequencies above 0.1 c/deg, the spectra demonstrate the expected "1/f" format, which is characteristic of most natural scenes[49]. The power law regression curves give:

$$c_s^2 = Au^{-B} \tag{4}$$

where A equals 0.0059 (for low contrast), 0.0647 (for high contrast); and B equals 0.807 (for low contrast) and 0.847 (for high contrast). The values of B are within the expected range for natural scenes[49].

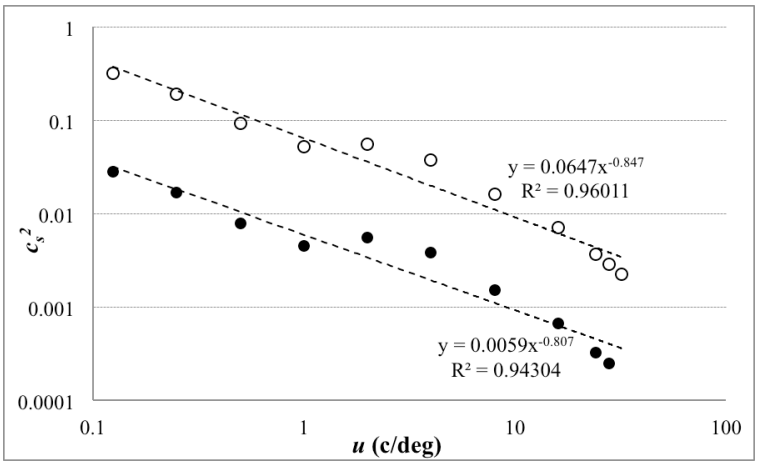


Figure 3. Contrast spectra for the low (black circles) and high contrast (white circles) versions of the Gallery image. Brolen lines represent the power low regression lines

### 3.3 cCSF

The data points in Figure 4 produce direct measurements of cCSF. The curves are theoretical predictions of cCSF and were obtained using Equation 3 with $c_i$ given from equation 1, and $c_s$ from equation 4. The constant in Equation 3 is theoretically independent of image contrast and, to a good approximation; this was found to be the case (0.005 for the low contrast image and 0.004 for the high contrast image).

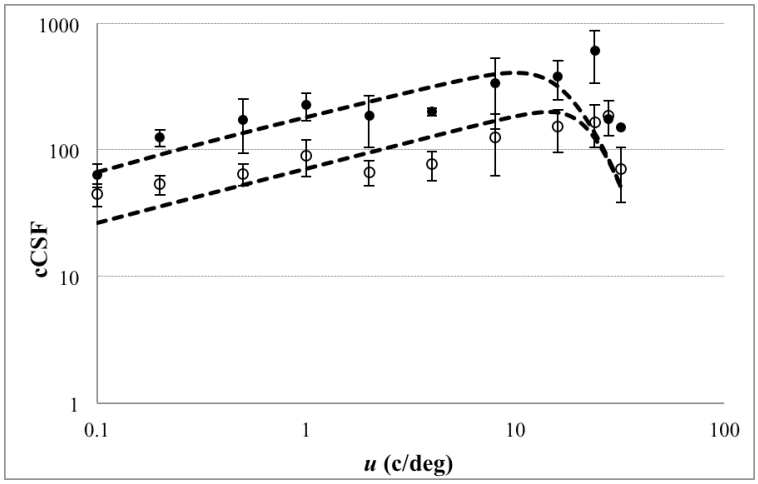


Figure 4. Measured (data points) and modeled cCSF (broken lines). Black circles: low contrast image, white circles: high contrast image.

In Figure 4 the cCSF is relatively flat in profile, compared with the bandpass nature displayed by the iCSF. The cCSF finally reduces in magnitude at high frequency. The high contrast cCSF is also lower in magnitude compared with the iCSF. This is clearly due to the increased masking impacting on band detection, which originates from other signal information in the high contrast image. The relationship between model predictions and measured data is very good.

### 3.4 iVPF

Data points in Figure 5 correspond to the measured iVPF. The curves are theoretical values calculated from equation 2. In the determination of these curves, appropriate values of $c_t$ (equal to $c_i$ for isolated band sensitivity) were obtained from Equation 3, for both low and high contrast images.

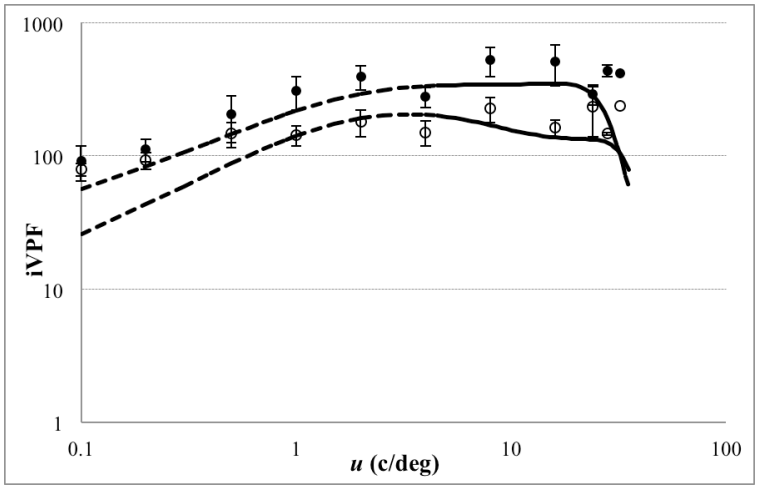


Figure.5. Measured (data points) and modeled iVPF (broken lines). Black circles: low contrast image, white circles: high contrast image.The curves are calculated from equations 1, 2 and 3 for both image contrast levels.

Good agreement exists between measured and modelled iVPF for the low contrast image. With the high contrast image, good agreement between measured and modelled values again exists for spatial frequencies at and above 1.0 c/deg. At lower frequencies, the model undershoots the measured data, which indicates a sensitivity close to the low contrast image. This deviation between model and measurement is discussed in section 4.6.

### 3.5 cVPF

cVPF results are given in Figure 6. Curves and data points are similarly defined as those in Figure 4. For the application of Equation 2, values of $c_t$ (equal to $c_c$ for contextual band sensitivity) were obtained from Equation 3, for both low and high contrast images. The deviation between measurement and model prediction is again evident at low spatial frequencies for the high contrast image.

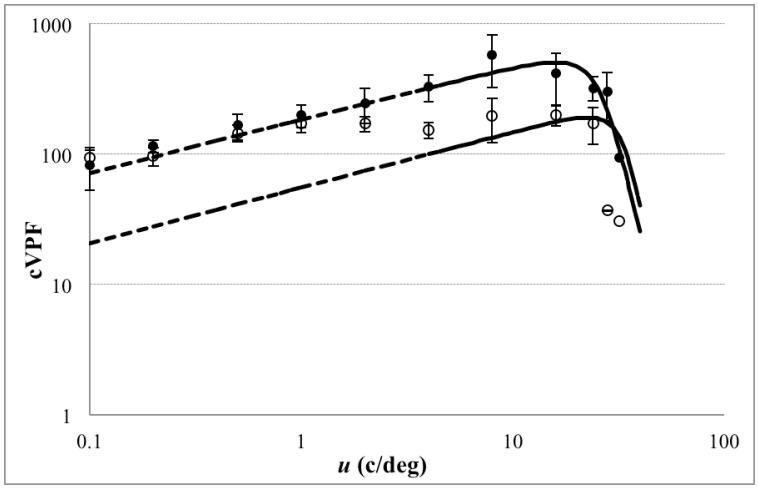


Figure 6. Measured (data points) and modeled cVPF (broken lines). Black circles: low contrast image, white circles: high contrast image.

### 3.6 Discussion on CSF and VPF results

As previously reported[2] the iCSF closely follows the profile of the standard sinewave grating CSF, appropriate for the size and luminance of the pictorial stimulus used in this study. Use of Equations 1, 3, and 4 enables good predictions of the cCSF to be made and this applies to both the low and high contrast versions of the image. Moreover, the results show that band detection is lowered as image contrast is raised. The overall profile of the cCSF reveals a slow increase in

contextual detection until a turning point is reached around 20 c/deg, due to optical limitations of the eye. If, as pictorial content varies the iCSF remains invariant and may be represented with the standard sinewave CSF relevant to both picture size and mean luminance, the model shows that the cCSF can be readily predicted, given the image's contrast spectrum. The generality of this finding will be tested with planned variations in pictorial content.

The iVPF obtained in this study shows a profile similar to cCSF. Agreement between measurements and the model is found at all frequencies in the low contrast image. Also, as Figure 5 illustrates, raising picture contrast reduces contrast discrimination for frequencies above about 1.0 c/deg, as predicted from the model. This behavior would in fact be expected from the so-called "dipper" function, which relates contrast discrimination to pedestal (or in our case band) contrast. Such functions are defined when the contrast difference $| c - c_s |$ is plotted against $c_s$. Typically the dipper shows that as pedestal contrast increases from zero (where discrimination is equivalent to detection), contrast discrimination sensitivity first *increases* to pedestal values of around 0.01, and then decreases approximately in accordance to Weber's law. The band contrasts existing in both the high and low contrast images used in our study are well above 0.01 in value.

The enhanced measured values of the iVPF at low frequencies in the high contrast image are unexpected and reveal visual discrimination sensitivity levels similar to those for the low contrast image. The model in its current form does not predict this behavior as seen in Figure 5. However, some experimental studies of the dipper function have revealed a reduction in slope and slight downward curvature of the nominal Weber region at low spatial frequency[50-52]. This demonstrates that at low frequency and relatively high contrast levels, discrimination sensitivity is significantly higher than would otherwise be expected. It is likely that this phenomenon is being reflected in our measured data. Many existing models of discrimination (including that represented in our model by Equation 2) do not address this particular low frequency dependent behavior. The visual mechanisms responsible for this discrimination finding are unclear. One possibility is that the compressive non-linear transduction mechanism outlined by Wilson[53] and others, which strongly influences discrimination of high contrast suprathreshold signals may, itself, be frequency dependent. This transduction mechanism is incorporated in the VPF model given in Equation 2, but not in a frequency dependent form. Further investigation of this particular non-linear feature will be conducted during current and future planned studies of the VPF model construct.

The cVPF results again show agreement between measured and modeled values for the low contrast image. The measured and modeled values obtained from the high contrast image show a similar deviation at low frequencies to that displayed by the iVPF results. Clearly, the low frequency mechanism operating to compensate for any reduction in discrimination sensitivity at high contrast, relates, at least in part, to the intrinsic perception of suprathreshold sinewave signals, whether defined from grating, or isolated band stimuli, i.e. it is not an effect relating to contextual viewing. However, it is very likely that other factors play a role in contextual band discrimination. For example, the perceived contrast of an individual band within a picture may not be fully reflected by the objective band contrast value[34]. Again, the existence of surround and overlaying spatial signals has been shown to directly lower perceived contrast of a suprathreshold test grating[54,55]. Interestingly, this effect appears greatest at frequencies below 1 c/deg[54]. At the moment, these additional perceptual considerations are not included in our modeling. These issues are important, particularly it would seem, at low frequencies, and we intend to address them in future work.

It is of interest to note that Haun and Peli[56] report that, under experimental conditions where random modifications in band contrast are performed, overall global pictorial contrast is determined primarily within the 1.0-6.0 c/deg frequency range. These authors suggest that contrast perception in a real scene should be weighted by a narrow band function centered around 2.0 c/deg and challenge the concept of contrast gain control. Such spatial frequency tuning of contrast discrimination is not observed in our cVPF measurements. The relatively flat profile of the cVPF is consistent with previous experimental investigations of band contrast discrimination in real scenes, when this has been directly measured. For example, the dipper functions obtained by Bex, Mareschal and Dakin[57] show that at a given (contextual) band contrast, contrast discrimination is similar in value at 1.0, 2.0 and 4 c/deg. The band contrast matching experiment performed by Bex and Makous[58] for a range of spatial frequencies (0.1-10.0 c/deg) also leads to the expectation of a relatively flat cVPF.

From a general consideration of factors influencing the overall subjective image quality of a reproduced pictorial scene, both contextual contrast detection (cCSF) and contextual contrast discrimination (cVPF) are of primary importance. Figure 7 shows a comparison between cCSF and cVPF for both the low and high contrast images.

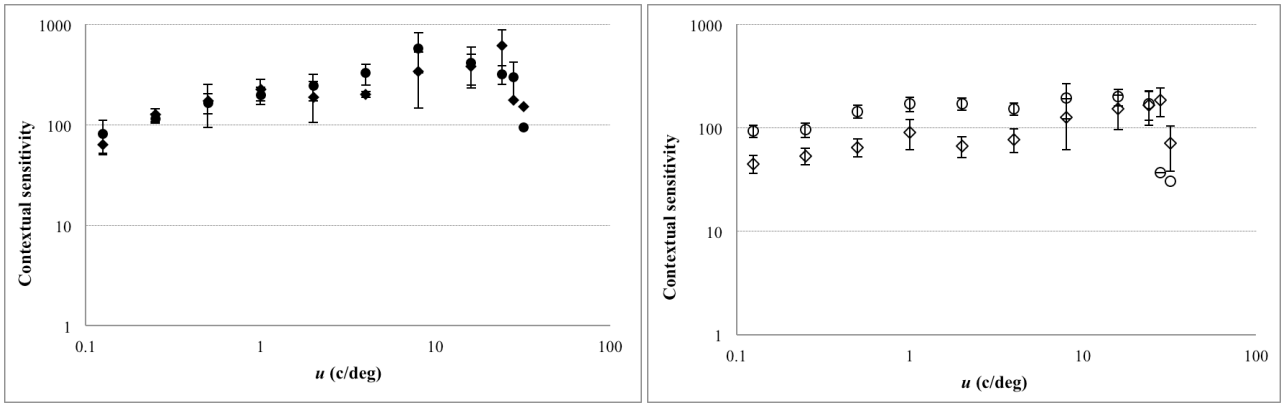


Figure 7. Comparison between measured cCSF (diamonds) and cVPF (circles) for the low contrast image (left) and high contrast image (right).

A feature of the data shown for the low contrast image is that both measured cCSF and cVPF are almost identical, and characterized by a gradual increase up to the optical limit of the eye. Similar conclusions can be drawn from the high contrast comparison, although detection is slightly reduced compared with discrimination. Overall gradient of the sensitivity responses is again low in this case. Given that the CSF is used as a visual weighting function in image quality models, (usually normalized, so that the absolute sensitivity magnitudes shown in Figure 7 have no real bearing), the results suggest that near equal weighting should be given to all frequencies, up to the limit of human spatial vision (acuity point). This weighting potentially incorporates visual masking effects, which are known to flatten the CSF[45]. We are continuing to measure both detection and discrimination sensitivities for a range of images of varying pictorial content, as described in reference 2. If similar results to our reported study were obtained, then our findings would significantly challenge current implementations of a standard bandpass (tuned to 1-2 c/deg) sinewave CSF in image quality modeling.

## 5. CONCLUSIONS

In this paper we described recent initial findings from on going research, undertaken to determine experimentally and further model threshold and suprethreshold (discrimination) contrast sensitivity functions from complex pictorial imagery. Results derived using one stimulus image at two different contrast levels (i.e. one stimulus at optimum visual contrast and another at considerably lower contrast than the first) indicate the following: Spatial contrast detection sensitivity in a real image (i.e. see cCSF) is characterized by a frequency response function, which is relatively flat up to around 20 c/deg (unlike the bandpass iCSF, which has a similar profile to a sinewave CSF for the given viewing conditions). Beyond this frequency, detection sensitivity decreases due to optical limitations of the eye. Spatial contrast discrimination sensitivity in a real image (see cVPF) displays a similar frequency response profile and magnitude to the detection function. This is a preliminary result, but if it holds for different types of imagery and viewing conditions, it will have implications in the way CSFs are considered, and possibly implemented, in image quality modeling. By experimenting with different contrast levels we found that increasing overall contrast of the image lowers both detection and discrimination sensitivities for frequencies above 1.0 c/deg. The results suggest that spatial detectors in the HVS attempt to normalize out contrast differences existing in natural scenes. Both detection and discrimination can be theoretically described using the Barten detection model and our extension to the Barten discrimination model.

## REFERENCES

[1] Triantaphillidou S, Jarvis J. and Gupta G., "Contrast sensitivity and discrimination of complex scenes", Proc. SPIE 8653 (2013).

[2] Triantaphillidou S, Jarvis J. and Gupta G., "Defining human contrast sensitivity and discrimination from complex imagery", Proc. SPIE 8901 (2013).
[3] Barten, P. J. G., [Contrast sensitivity of the human eye and its effects on image quality], SPIE Press, Washington, (1999).
[4] Nagaraga, N., S. "Effective luminance noise on contrast thresholds", JOSA, 54, 950-955 (1964).
[5] Barten, P. J. G., [Contrast Sensitivity of the Human Eye and its Effects on Image Quality], SPIE Press, Washington, chapter 7 (1999).
[6] Granger, E. M. and Cupery, K. N., "An optical merit function (SQF) which correlates with subjective image judgments", Photograph. Sci. Engng, 16(3), 221-30 (1972).
[7] Snyder, H. L., "Image quality and observer performance", in [Perception of displayed information] (Biberman, L.M ed.), Plenum Press, New York and London, chapter 3 (1973)
[8] Barten, P. J. G., [Contrast sensitivity of the human eye and its effects on image quality], SPIE Press, Washington, chapter 8 (1999).
[9] Töpfer, K and Jcobson, R.E., "The relationship between objective and subjective image quality", J. Inf. Rec. Mats., 21, 5-27 (1993).
[10] Jenkin, R. B., Triantaphillidou, S. and Richardson, M. A., "Effective pictorial information capacity as an image quality metric", Proc. SPIE 6494, ( 2007).
[11] Schade, O. H., Optical and photoelectric analog of the eye. J. Opt. S. Am., 46, 721-39 (1956).
[12] Jacobson, R.E and Triantaphillidou, S., "Metric approaches to image quality", in [Colour Image Science: Exploiting Digital Media], (MacDonald, W. L. and Luo M. R. ed.), John Wiley and Sons, London, UK, Chapter 18 (2002),
[13] Triantaphillidou, S., Allen, E., and Jacobson, R.E., "Image Quality Comparison Between JPEG and JPEG2000. II. Scene Dependency, Scene Analysis, and Classification", J. Im.Sci. Techn., 51, 259-270 (2007).
[14] Oh, K. H., Triantaphillidou, S. and Jacobson, R. E., Scene classification with respect to image quality measurements", Proc. SPIE 7529 (2010).
[15] Haun A. and Peli E., "Is image quality a function of contrast perception?", Proc. of SPIE 8651, (2013)
[16] Daly, S., "The visible difference predictor: an algorithm for the assessment of image fidelity", in [Digital Images and Human Vision] (Watson, A. B. ed.) 179-206, The MIT Press, Cambridge MA (1993).
[17] Jacobson, R. E, Ford, A. M. and Attridge, G. G., "Evaluation of the effects of compression on the quality of images on a soft display", Proc. SPIE 3016 (1997).
[18] Farrell, J. E., "Image Quality Evaluation", in [Color Imaging: Vision and Technology], (MacDonald, W. L. and Luo M. R. ed.), John Wiley and Sons, London, UK, chapter 15 (1999),
[19] Lubin, J., "The use of psychophysical data and models in the analysis of display system performance", in [Digital Images and Human Vision] (Watson, A. B. ed.) 163-178 The MIT Press, Cambridge MA (1993).
[20] Ahumada, Jr., A. J., "Computational image quality metrics a review", Proc. SID 24, 305-308 (1993).
[21] Silverstein, D. A. and Farrell E.J., "The Relationship Between Image Fidelity and Image Quality", Proc. IEEE International Conference on Image Processing I (1996).
[22] Zhang, X., Silverstein, D. A., Farrell, J. E., and Wandell, B., "Color Image Quality Metric S-CIELAB and its Application on Halftone Texture Visibility", Proc. IEEE COMPCON (1997).
[23] Fairchild, M. D. and Johnson, G. M., "The iCAM framework for image appearance, image differences, and image quality", J. Electron. Imaging, 13 (2004).
[24] Orfanidou, M., Triantaphillidou S. and Allen, E., "Predicting image quality using a modular image difference model", Proc. SPIE 6808 (2008).
[25] Rohaly, A. M., Ahumada, Jr., A. J. and Watson, A. B., "A comparison of image quality models and metrics predicting object detection", SID 26, 45-48 (1995).
[26] Rogowitz, B. E., Papas, T. N. and Allebach, J. P., "Human vision and electronic imaging", J. Electronic Im., 10, 10–19 (2001).
[27] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli E. P., "Image quality assessment: From error visibility to structural similarity", IEEE Trans Image Process 13, (2004).
[28] Sheikh, H.R., Bovik, A.C. and deVeciana, G., "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Process., 14, 2117–2128 (2005).
[29] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans Image Process, 15, 430–444, (2006).

[30] Chandler, D.M. and Hemami, S. S., "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", IEEE Trans. Image Process, 16 (2007).
[31] Haun. A and Peli E., "The complexities of complex contrast", Proc. SPIE 8292 (2012).
[32] Peli E., "Contrast sensitivity function and image discrimination,", J. Opt. Soc. A. A,18, 283-293 (2001).
[33] Haun A and Peli E., Measuring the perceived contrast of natural images" SID Digest 302-304 (2011).
[34] Carney, T., Tyler, C.W., Watson, A.B., Makous, W., Beutter, B., Chen, C.-C., Norcia, A.M., and Klein, S.A., "Modelfest: year one results and plans for future years," in Proc. SPIE 3959, 140–151 (2000).
[35] Watson, A.B., and Ahumada, A.J., "A standard model for foveal detection of spatial contrast," J. Vision 5, 717–740 (2005).
[36] Rovamo, J., Mustonen, J. and Nasanen, R. "Modelling contrast sensitivity as a function of retinal illuminance and grating area", Vision Res. 34, 1301-1314 (1994)
[37] Gobbi, P. G., [Modeling the Optical and Visual Performance of the Human Eye], SPIE Press, Bellingham, MA (2103).
[38] Jarvis, J.R. and Wathes, C.M. "Mechanistic modelling of vertebrate spatial contrast sensitivity and acuity at low luminance", Visual Neurosci, 29, 169-181 (2012).
[39] Crozier, W. J., "On the variability of critical illumination for flicker fusion and intensity discrimination", J. Gen. I. Psychol., 19, 503-522 (1935).
[40] Roufs, J. A., "Dynamic properties of vision - V1. Stochastic threshold fluctuations and their effect on flash-to-flicker ratio", Vision Res., 14, 871-888 (1974).
[41] McAnany, J. J and Alexander, K. R., "Spatial contrast sensitivity in dynamic and static additive noise", Vision Res., 50, 1957-1965 (2010).
[42] Legge, G. E., Kersten, D. and Burgess, A.E., "Contrast discrimination in noise", J. Opt. Soc.Amer., A, 4, 391-404 (1987).
[43] Pelli, D.G., "Uncertainty explains many aspects of visual contrast detection and discrimination", J. Opt. Soc. Am. A, 2, 1508-1532 (1985).
[44] Stromeyer, C.F. and Julesz, B., "Spatial frequency masking in vision: critical bands and spread of masking", J. Opt. Soc. Am. A, 62, 1221-1232 (1972).
[45] Van Meeteren, A. and Valeton, J., "Effects of pictorial noise interfering with visual detection", J. Opt. Soc. Amer., A, 5, 438-444 (1988).
[46] Blackwell, K.T., "The effect of white and filtered noise on contrast detection thresholds", Vision Res., 38, 267-280 (1998).
[47] McAnany, J.J and Alexander, K.R., "Spatial contrast sensitivity in dynamic and static additive noise", Vision Res., 50, 1957-1965 (2010).
[48] Kingdom, F. A. A. and Prins, N., [Phsychophysics], Academic Press: Elsevier, London (2010).
[49] Billock, V.A.,"Neural acclimation to1/f spatial frequency spectra in natural images transduced by the human visual system", Physica D, 137, 379-391(2000).
[50] Legge, G. E., "Spatial frequency masking in human vision: Binocular interactions", J. Opt. Soc. Am., 69, 838-847 (1979).
[51] Kulikowski, J. J., "Effective contrast constancy and linearity of contrast sensation", Vision Res., 16, 1419-1431 (1976).
[52] Yang, J. and Makous, W. "Modeling pedestal experiments with amplitude instead of contrast", Vision Res., 35, 1979-1989 (1995).
[53] Wilson, H. R., "A transducer function for threshold and suprathreshold human vision", Biol. Cybern., 38, 171-178 (1980).
[54] Meese, T. S. and Hess, R. F. (2004), " Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin", J. Vision, 4, 843-859 (2004).
[55] Xing, J and Heeger, D.J. "Measurement and modeling of center-surround suppression and enhancement", Vision Res., 41, 571-583, (2001).
[56] Haun, A.M. and Peli, E. "Perceived contrast in complex images", J. Vision, 13, 1-21 (2013).
[57] Bex, P.J., Mareschal, I and Dakin, S.C. "Contrast gain control in natural scenes", J Vision, 7, 1-12 (2007).
[58] Bex, P.J. and Makous, W. " Spatial frequency, phase and the contrast of natural images", J.Opt. Soc.Am., A, 19, 1096-1106 (2002).

# APPENDIX

## The theoretical relationship between contextual detection (cCSF) and isolated detection (iCSF) through application of the linear amplification model (LAM)

The impact of spatiotemporal noise on signal contrast detection can be quantified through the *linear amplification model* (LAM)[3,42]. This model has been verified experimentally through a number of studies employing sine-wave gratings embedded in both static and dynamic noise[3,42,45,47]. If the signal and noise cover the same stimulus area, the LAM states that for a given spatial frequency:

$$(\Phi)_c - (\Phi)_i = \lambda^2 (\Phi)_n \quad \text{(A1)}$$

where $(\Phi)_c$ and $(\Phi)_i$ denote the power spectral density of the signal at the *threshold* of detection either with or without spatiotemporal noise respectively. $(\Phi)_n$ is the power spectral density of the noise. The constant $\lambda^2$ is a dimensionless number, independent of spatial or temporal frequency. It relates to the reciprocal of *sampling efficiency,* which describes the ability of an observer to make use of stimulus information relative to an ideal observer[47]. Barten[3] also indicates that this constant can also be described in terms of the Crozier coefficient. In other words, $\lambda^2$ is essentially a basic visual constant.

As now shown, our analysis treats individual signal bands as also acting as sources of noise operating on other bands in the image.

The general relationship between the two-sided version of Φ and RMS contrast *c* for static images is then given by:

$$(c)^2 = \iint \Phi(u,v) 2\,du 2\,dv \quad \text{(A2)}$$

where the limits of each integration are 0 and ∞, and the terms *u,* and *v* denote spatial frequency.

For each filtered band in our experiments we define *c* as:

$$(c)^2 = \frac{1}{N} \sum ((L_j - L_{mean}) / L_{mean})^2 \quad \text{(A3)}$$

where $L_j$ is the luminance of the $j^{th}$ pixel in the band, $L_{mean}$ is the mean luminance of the image and *N* is the number of pixels. The summation shown in Equation A3 is conducted over the range *j*=1.0 to *N.* Note that for all bands considered in our examination, *N* is constant and equal to the total number of pixels in the image. Our particular band signals were produced using rotationally symmetric log cosine filters[1], which define *u*=*v*. In this analysis, each signal band is mathematically approximated by an idealized rectangular spectrum of 1 octave width, defined by a frequency difference of *(a - b)* c/deg. A constant spectral density is also assumed within a given band. For such a band we have from Equation A2:

$$(c)^2 = (2(a-b))^2 \Phi \quad \text{(A4)}$$

Noise masking effects on grating detection at a given frequency *u* c/deg emanate mainly from spatial frequencies occurring within +1 and -1 octave of this frequency[44]. We assume that the same occurs for masking of individual signal bands in our image. In our experimentation, we initially measure band contrast sensitivity when presented in isolation to the observer i.e. all other image bands removed. The measured sensitivity therefore represents a response to a combination of all frequency components defined within the octave. We then measure the band sensitivity with all the other signal bands present, i.e. within the full context of the pictorial image. Any change in detection sensitivity compared with the isolated condition, will reflect the impact of spatial frequencies *outside* of the signal octave.

The impact of masking from a noise source of frequency $u_n$, on a signal at frequency *u*, can be quantified through:

$$\Phi_n(u) = \int_0^\infty \psi(u_n, u)\, \Phi(u_n) / u\, du \quad \text{(A5)}$$

where $\Phi(u_n)$ represents the spectral density of the noise source at frequency $u_n$, remote from the signal band. $\Phi_n(u)$ is the *effective* spectral density of this noise operating at frequency $u$.

The function $\psi$ found empirically by Barten[3] from a study of the Stromeyer and Julesz[44] data is given by:

$$\psi(u_n,u) = 0.747 \exp(2.2 \ln^2(u_n/u)) \quad \text{(A6)}$$

Equation A6 represents a log normal function described by a Gaussian distribution of $\ln(u_n/u)$, with a half-height width of two octaves, consistent with the premise that noise masking on a given signal band will be mainly generated from spatial frequencies within the two flanking signal bands. Consider now, a band signal centered on frequency $u$, (with squared RMS contrast denoted by $c_s^2$ ). This represents one point on the contrast spectrum. For the specific image from which we have measured detection sensitivity (low contrast Gallery) we find that, to a good first approximation:

$$(c(u))_s^2 = 0.0056u^{-0.778} \quad \text{(A7)}$$

If it assumed that the noise at $u$ emanates from the two flanking signal bands (centered at +1 and -1 octaves from $u$), then for each of these Equation 5 can be stated as:

$$\left[\Phi_n(u)\right]_{-1} = \Phi_{-1}\int_a^b \psi(u_n,u)\,/u\,du \quad \text{(A8)}$$

for frequencies in the signal band -1 octave below the band at $u$ and:

$$\left[\Phi_n(u)\right]_{+1} = \Phi_{+1}\int_a^b \psi(u_n,u)\,/u\,du \quad \text{(A9)}$$

for frequencies +1 octave above.

The integration limits $(a,b)$ are $(1/\sqrt{2}u,\ 1/2\sqrt{2}u)$ for the -1 octave flanking band and $(2\sqrt{2}u,\ \sqrt{2}u)$ for the +1 octave flanking band.

Using Equation A4, Equations A8 and A9 can be re-written as Equations A10 and A11:

$$\left[\Phi_n(u)\right]_{-1} = \left(c_s^2\right)_{-1}\left(2(a-b)\right)^{-2}\int_a^b \psi(u_n,u)\,/u\,du \quad \text{(A10)}$$

$$\left[\Phi_n(u)\right]_{+1} = \left(c_s^2\right)_{+1}\left(2(a-b)\right)^{-2}\int_a^b \psi(u_n,u)\,/u\,du \quad \text{(A11)}$$

The total noise spectral density at $u$ is then given by:

$$[\Phi_n(u)] = [\Phi_n(u)]_{-1} + [\Phi_n(u)]_{+1} \quad \text{(A12)}$$

Equation 12 was evaluated for each signal band as its central frequency $u$ varied. The integrals were calculated using Simpsons rule. It was found that, at any given $u$:

$$\Phi_n = 0.8\,\Phi_s \quad \text{(A13)}$$

where $\Phi_s$ denotes the spectral density of the signal band at $u$.

Combining Equation A1 (the LAM) with Equations A4 and A13 gives the prediction that, at a given value of $u$:

$$[c_c^2] - [c_i^2] = \lambda^2\, 0.8\, c_s^2 \quad \text{(A14)}$$

The terms $[c_c^2]$ and $[c_i^2]$ denote the *contextual* and *isolated* detection threshold values (squared) for RMS contrast at the band at $u$. Thus, the algebraic difference between these two quantities is directly proportional to the square of the contrast spectrum at $u$.