Testing the Agreement between Contrast Discrimination and Appearance Measured with Scaling Methods

Burak Kiran

March 27, 2023

Technische Universität Berlin Fakultät IV
Elektrotechnik und Informatik Institut
für Technische Informatik und Mikroelektronik
Department Computational Psychology

Thesis for obtaining the academic degree of
Bachelor of Science in the study of Computer Science.

First Examiner: Dr. Guillermo Aguilar
Second Examiner: Prof. Dr. Felix Wichmann

I hereby declare that the thesis submitted is my own, unaided work, completed without any unpermitted external help. Only the sources and resources listed were used.

Berlin,

____

Burak Kiran

Abstract

In the psychophysics literature it is still an open question whether experiments based on performance and based on appearance are equivalent with each other. Recently Shooner and Mullen (2022) investigated this relationship in the domain of contrast perception, claiming that discrimination can predict appearance. This thesis investigated this relationship by creating discrimination scales from previous contrast discrimination data and comparing them to the previous appearance experiment results. The comparison to scales from literature showed that the simulated discrimination scales had a large difference in shapes and precisely at 16 cpd this difference was significant. The results support the idea that appearance and performance are distinct type of measurements and may rely on distinct processes. Furthermore during the investigation an inverse correlation between detection thresholds and perceived contrast at high contrast was found.

Zusammenfassung

In der psychophysikalischen Literatur ist es noch offen, ob Experimente auf der Grundlage von Leistung und auf der Grundlage des Aussehens einander gleichwertig sind. Shooner and Mullen (2022) untersuchte diese Beziehung im Bereich der Kontrastwahrnehmung und behauptete, dass Diskriminierung das Aussehen vorhersagen kann. Diese Thesis untersuchte diese Beziehung, indem sie Diskriminationsskalen aus früheren Kontrastdiskriminationsdaten erstellte und sie mit den Ergebnissen früherer Erscheinungsexperimente verglich. Der Vergleich mit Skalen aus der Literatur zeigte, dass die simulierten Diskriminationsskalen einen großen Unterschied in der Form hatten und gerade bei 16 cpd dieser Unterschied signifikant war. Die Ergebnisse unterstützen die Idee, dass Aussehen und Leistung unterschiedliche Arten von Messungen sind und auf unterschiedlichen Prozessen beruhen können. Weiterhin wurde bei der Untersuchung eine inverse Korrelation zwischen Detektionsschwellen und empfundenem Kontrast bei hohem Kontrast festgestellt.

Acknowledgement: I wish to thank my supervisor Dr. Guillermo Aguilar for all the help he provided through finishing my thesis.

Contents

1 Introduction

Naively we tend to think that we perceive the world objectively, uniquely, and as it is. However, an example striking of perception differences is two people arguing about the color of a dress. In 2015 a color of a dress was discussed heavily through social media. People had different opinions such as the dress being white-gold or black-blue. The perception of the store owner that sold the dress was that the dress was produced in black-blue colors [1]. "Why are people seeing this dress differently" was the question asked by many researchers. Aston and Hurlbert (2017) believe that people's perception of the dress was influenced by a great number of parameters such as device brand-model, angle, room lighting, and many others. The work done by Lafer-Sousa, Hermann, and Conway (2015) and many others showed that the difference in the people's answers lay in the perception of color. This example shows that perception can be variable among individuals and that their variability is worth studying.

Figure 1: **Picture of a dress, where people argued about it's color. Popular answers to the asked color were white-gold or black-blue.**

Differences in perception are not limited to color or even to the sense of seeing. In terms of object weight perception, a commonly witnessed example is the same object being perceived as "heavier" or "lighter" according to the latest lifted object weight. A person, who lifted 5 kg before lifting a 20 kg object, would perceive as 20 kg heavier and may perceive the weight of the object as 25 kg. The same person, who lifted 40 kg before lifting 20 kg object, would perceive a 20 kg as lighter and may perceive the weight of the object as 15 kg.

Discrimination ability is another aspect where people differentiate among themselves. One real-life example is the weight of the objects. The weight difference between 15 and 16 kg is harder to discriminate than the difference between 1 and 2 kilos even though

[1] https://en.wikipedia.org/wiki/The$_{d}ress/media/File : The_{d}ress_{b}lueblackwhitegold.jpg$

they both differ by 1 kilo. The same absolute difference, 1 kilo, in one case is harder to discriminate than in the other because our ability to discriminate does not depend on the "relative" difference, relative to the absolute value (Weber's Law). Weber published that our perception of weight differences (Δw) depends linearly on the absolute weight reference w. The published relationship $\Delta w = kw$ is now known as "Weber's Law" and established the beginning of the scientific study of perception: psychophysics.

1.1 Psychophysics

Psychophysics surfaced historically around the 1800s. Khaleefa and other authors (1999) believe Gustav Theodor Fechner is the founder of psychophysics. Fechner (1860) defined psychophysics as the science, which investigates the perception of organisms when encountering a stimulus in an environment. A stimulus can be seen as any object that can be identified with any five senses of an organism. Every stimulus has different physical attributes that identify the stimulus with a certain unit type such as a meter as a measurement of height or kg as a measurement of weight. The physical attributes of stimuli are objective and can be measured with physical instruments. Having a second opinion about the physical attribute values is not possible, because they are measured with tools such as rulers to measure the height that is certain. This is not the case with perception. Perception is subjective, it can differ among people and is the main focus of psychophysics. Perceptual values are measured with experiments, there are no tools to measure the perceptual values. Perceptual variables such as perceived weight and perceived height are measured differently for each human.

Subjectivity in audition and vision are important topics in psychophysics. This thesis focuses on vision. Contrast, spatial frequency, flicker frequency, and color are some of the heavily researched topics in visual psychophysics. Physics defines these attributes through precise units and the science of psychophysics aims to find the effect of these attributes in our vision. "Do we perceive this stimulus as high contrasted, do we perceive this stimulus as high flickered?" are some example questions that are researched in psychophysics.

1.2 Classifications in Psychophysics

There are different classifications for experiments done in psychophysics. Prins et al. (2016) divide the experiments into eight groups including "Performance versus Appearance" and "Threshold versus Suprathreshold". The authors claim that the goal of this grouping is to categorize the design and possible outcomes of the different experiments. Furthermore, it is stated that these groupings differ from each other and it is best to handle them differently, without assuming they have a connection. This statement is not accepted as a fact by the scientific community but is believed to be true by the majority.

This thesis focuses on the classification “Performance versus Appearance”.

1.2.1 Performance versus Appearance

Prins et al. (2016) define performance in experiments as the measurement of aptitude to detect a stimulus on to discriminate between two or more stimuli. These experiments measure the correctness rates of the observers. The performance experiments can be compared with the commonly used multiple-choice tests in the way that there is always one correct answer. In psychophysics experiments, the stimuli are shown multiple times to increase the quality of the results. In performance experiments, it is used to determine the just noticeable difference as well, which is where observers give the correct answer above a certain amount of times. This value is commonly fixed at 75% and may vary from one experiment to another experiment. In performance experiments, the measurement is called a threshold. Thresholds are defined as the physical values, where observers gave the right answer to the question asked by the experimenter at the determined JND percentage (usually 75%). The region where the physical variable is above thresholds is called the suprathreshold region.

Performance experiments are an example of discrimination experiments, where observers try to discriminate a stimulus from a fixed stimuli determined by the experimenter. For example, asking observers the question “Which line is more straight?” and showing Figure 2. The line below is fixed and the curvature of the line above is changed through the experiment. The goal of this experiment is to measure the last points, where observers can recognize that two lines differ from each other and these points are called the thresholds. Since one line is more straight than the other one, in this figure the bottom line, the observers have a right and wrong answer and their performance (how much of the time presented, the observer gave the right answer) is measured by the experimenter. The experiment is continued until a threshold for each fixed line is found. The observers commonly give different answers to the stimuli combinations, because of the internal noise of the observers.

Figure 2: **An example of a performance experiment. In this experiment, the curvature of the upper line varies and observers were asked: “Which of the two lines is straighter, the one above or the one below?” Adapted from Prins et al. (2016).**

Appearance is defined differently through different resources. Prins et al. (2016) characterize appearance in experiments as the "apparent magnitude of a stimulus dimension". In simple words, the appearance experiments measure how things appear, how they "look like". Observers answers given to the question of the experiment can not be right or wrong. The goal of these experiments is to measure the relationship between the perceptual variable and the physical variable itself and therefore the correctness of the observer is not relevant. The questions similar to the questions "which of the two stimuli appear... than the example stimuli?" or "which stimulus appears ..." are the questions these types of experiments investigate.

An example appearance experiment is shown in Figure 3. In this experiment observers were asked to choose the stimuli that appeared longer. The goal of this experiment is to measure the perception that our minds create to make one of these stimuli appear longer even though they have the same length. In other words, to see the relation between the physical variable length and the perceptual length that the human mind generates. As in all appearance experiments, there is no correct answer in this experiment, the experiment aims to create a scale that maps the perceived length to the physical length.

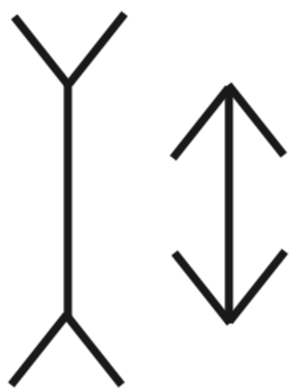


Figure 3: **The figure shows the famous Muller-Lyer illusion. In this experiment observers are asked: "Which of the two lines appears longer, the one on the right or the one on the left?". Even though the two central lines are equal in length, the left one appears longer than the right one. Adapted from Prins et al. (2016).**

As stated there is no agreement, on whether these two types of tasks measure the same underlying process. In other words, if performance can be used to predict appearance, and vice versa.

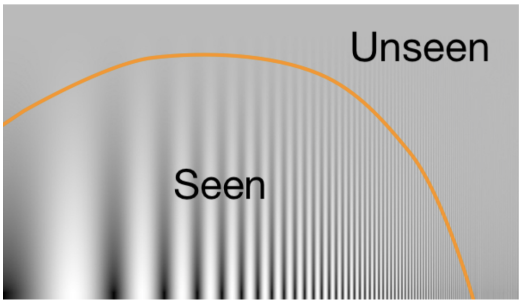
2 Previous Work on Contrast

Contrast detection, discrimination, and appearance are one of the most studied aspects of human vision. Contrast has a variety of accepted definitions in the scientific literature. According to Kukkonen, Rovamo, Tiippana, and Näsänen (1993) the best way to choose the right definition for contrast is, by fitting a different definition of the contrast according to the stimuli. The authors give examples of three stimulus types. For spot detection, the best definition is “Weber fraction or light energy”, for periodic stimuli the best definition is “Michelson contrast” and “RMS contrast” is the best definition for aperiodic complex stimuli. Subjectively, contrast is the average difference of luminance in pixels or in other words how different each pixel look on average.

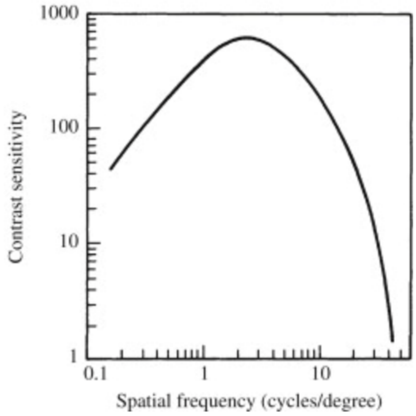
Contrast is one of the main topics of research in vision science and the focus of this thesis concerns the dichotomy of “Performance” and “Appearance” in psychophysics. In the following, some selected relevant work done on perceptual contrast and contrast thresholds from a wide historical period will be shortly reviewed.

2.1 Contrast sensitivity function Campbell and Robson (1968)

In 1956 Schade became one of the first people who mentioned “Contrast Sensitivity Function” and further on it has been investigated in many scientific researches. Commonly shortened as CSF, it represents the relationship between contrast sensitivity and spatial frequency. Figure 4a and Figure 4b show the commonly used versions of the CSF. The CSF version in Figure 4a shows the spatial frequency intervals, where humans can see and not see a certain contrast level. The CSF version in the Figure 4b shows that the contrast sensitivity is low when the spatial frequency is low (close to 0cpd). In mid-level spatial frequencies (around 4-5 cpd) contrast sensitivity increases and reaches its peak. After the peak, CSF starts decreasing towards 40 cpd and not varying significantly after 25 cpd.



(a) **Taken from understandinglowvision[2].**



(b) **Taken from Bovik (2009).**

Figure 4: **The figures show two representations of CSF. In both sub-figures, the x-axis shows increasing spatial frequency and the y-axis shows contrast sensitivity. In both graphs, the region below the line is where humans can see the contrast, and above the line contrast is unseen.**

To summarize both figures, when the spatial frequency is at the lowest or highest region, the visual system can not discriminate contrast as well as at mid-level spatial frequencies.

2.2 Georgeson and Sullivan Experiment (1975) and Contrast Constancy

"Contrast Constancy" refers to a phenomenon found by Georgeson and Sullivan (1975) which related contrast discrimination with contrast. The authors created two experiments to investigate contrast thresholds and perceived contrast at different spatial frequencies.

The first experiment (a matching experiment, a common appearance method) investigated how spatial frequency affected the perceived contrast. Observers were shown eight stimuli. The contrast of these stimuli varied between approximately 0.6 (high) and 0.003 (low). Observers were asked to adjust the contrast of a stimulus until it matched the contrast of the fixed reference stimulus. The spatial frequency of the stimulus varied from 0.25 cpd to 25 cpd. The empty dotted graphs in Figure 5 show the results of the experiments at different spatial frequencies. The full dots and squares in the middle of

[2] https://understandinglowvision.com/2-0-contrast-sensitivity-function/

the figure show the physical contrast values of each reference stimulus. In high contrast stimuli (contrast above 0.2), perceptual contrast is similar to physical contrast through all spatial frequencies. This shows that observers matched the reference stimulus at a similar contrast value when the stimuli had high contrast (0.6), producing lines that are flat in the graph (e.g. lowest in Figure 5).

In the low contrast part of the experiment, observers matched the reference stimulus well, only when the spatial frequency was in mid values (4-5 cpd). On lower or higher than 4-5 cpd spatial frequencies, observers adjusted the stimulus higher than the contrast of the reference stimulus. On very low contrast stimuli and lowest/highest spatial frequency, observers adjusted the contrast of the stimulus 10 times higher than the reference stimulus. This can be seen in Figure 5 as an inverted parabola for lower contrast, resembling the CSF. Furthermore, observers matched the same contrast for the stimuli between 0.003-0.1, when the spatial frequency was at 25 cpd. This matching was also seen at 0.25 cpd but slightly worse.

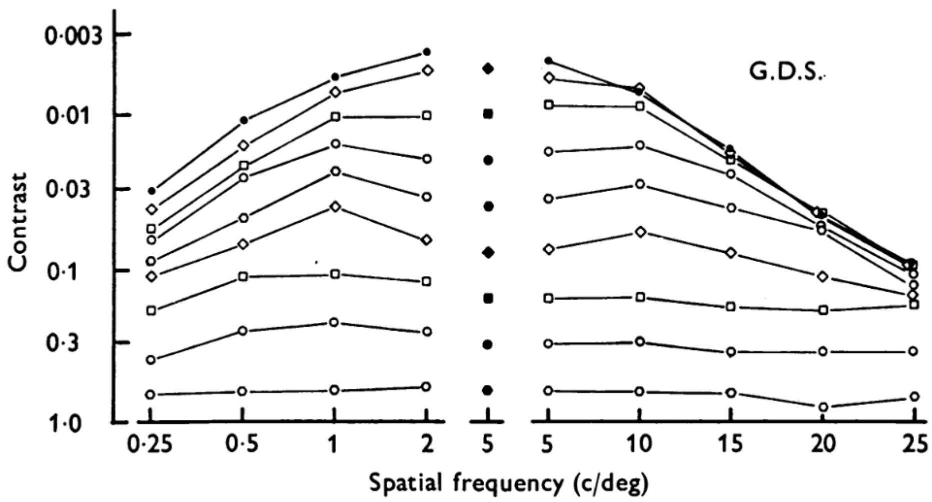


Figure 5: **The figure shows the relationship between spatial frequency and contrast. The full square and dots show the physical contrast values of each stimulus and the empty dotted graphs show the perceptual contrast of each stimulus on each spatial frequency. The full dotted graph shows the thresholds in each spatial frequency for low contrast. Taken and modified from Georgeson and Sullivan (1975) experiment.**

The second experiment was a discrimination experiment (performance method). The experiment was done only for low-contrast stimuli (around 0.003). Observers were asked to choose the stimuli with the highest contrast, and the effect of spatial frequency on contrast thresholds was measured in different spatial frequencies. The results (Figure

5) shown with filled circles show that in different spatial frequencies different thresholds are observed in this contrast region (0.003-0.03). The thresholds graph followed the traditional CSF behavior by thresholds being lower in mid-spatial frequency values than the higher or lower spatial frequency values. Thresholds reached the minimum contrast at 4-5 cpd and increased with distance to 4-5 cpd such as 25 cpd or 0.25 cpd.

To summarize, the phenomenon states that firstly for low contrast (below 0.1), the perceived contrast values are similar to physical contrast values, if the spatial frequency is in mid values (4-5 cpd). If the spatial frequency values are far from mid values (higher or lower than 4-5 cpd) then the perceived contrast increases and the difference between physical contrast and perceived contrast increases. Furthermore, at a very high spatial frequency (25 cpd) all stimuli between approximately 0.02-0.003 have the same perceptual contrast value. Secondly, for high contrast (above 0.3) stimuli, the spatial frequency did not affect the perception of the observers and perceived contrast was similar to physical contrast. Aside from the phenomenon, they have also found that the threshold function was similar to the known CSF.

2.3 Bird, Henning and Wichmann Experiment (2002)

Georgeson and Sullivan's experiment showed that two different regions exist in contrast perception: On low contrast, contrast detection, and discrimination depend on spatial frequency, at high contrast it does not. An interesting open question is thus to investigate the transition between contrast sensitivity and suprathreshold discrimination (Bird et al. (2002)) (and eventually appearance (Shooner and Mullen (2022))).

In 2002 Bird et al. conducted contrast discrimination experiments at suprathreshold levels for several spatial frequencies (0.00 cpd, 2.09 cpd, 4.19 cpd, 8.37 cpd, and 16.74 cpd). These spatial frequencies are the ones where the perceived contrast in Georgeson and Sullivan's experiment differed heavily. Observers were asked to choose the stimuli with the higher contrast. The authors ran the experiment with three different JNDs: 60%, 75%, and 90%. Figure 6 shows the normalized results of an observer and JND was set at 75%. The normalization was made on both axes according to the "best estimate of the contrast that corresponded to 75%" for each observer. The authors furthermore state that the results for 16.74 cpd were not reliable.

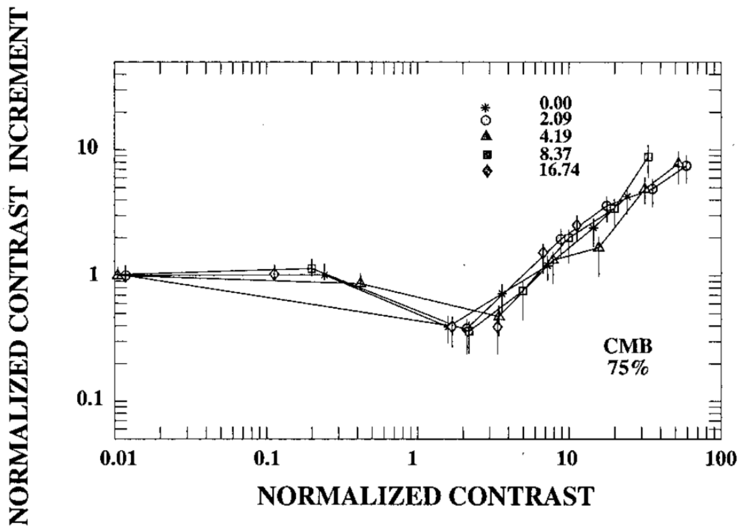
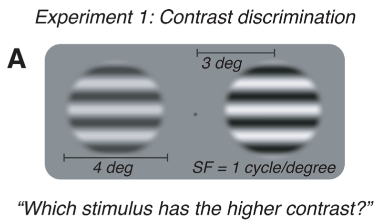


Figure 6: **The figure shows the contrast discrimination experiment results in different spatial frequencies of an observer. The x-axis shows the normalized contrast and the y-axis shows the normalized contrast increment. Taken from Bird et al. (2002)**

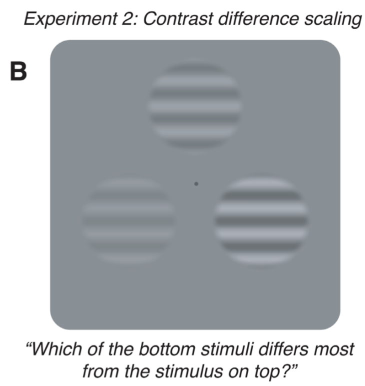
The normalized results (JND 75%) show firstly that for all spatial frequencies a "dipper function" is obtained. A common finding in contrast discrimination (Legge and Foley (1980)). The results show the so-called "dipper function", which moves toward the minimum threshold from the starting point and then increases after it reaches the minimum. This means the increment needed to notice the contrast difference between two stimuli decreases at very low contrast and then increases with the highest contrast. After the function reaches the minimum threshold, the increase is linear in log-log coordinates, meaning that the slope of that part of the graph is a constant. All graphs start by decreasing after their starting point and decrease until their minimum. After they have reached their minimum they increase in a linear form until their last point with varying slopes. The measured slopes for spatial frequencies were 0.92 for 0.00 cpd, 0.67 for 2.09 cpd, 1.00 for 4.19 cpd, 1.23 for 8.37 cpd, and 1.57 for 16.74 cpd. The authors find the slopes consistent with the experiment from Skottun, Bradley, Sclar, Ohzawa, and Freeman (1987) and disclosed from these two experiments that spatial frequency does not effect the slopes in the suprathreshold region. If discrimination at suprathreshold contrast is not different from different spatial frequencies, then it is still an open question how contrast constancy is achieved, given that the differential sensitivity (CSF) should be somehow "compensated" to have stimuli appearance look equal in contrast at high contrast values.

2.4 Shooner and Mullen Experiments (2022)

As stated earlier, it is still an open question whether performance and appearance are equivalent in the field of Psychophysics. Recently, Shooner and Mullen (2022) claimed that in the case of contrast perception a link between performance and appearance does exist. They designed an experiment to investigate that relationship. The goal of the experiment was to test whether scales derived from a contrast discrimination experiment (performance method) agree with the scales derived using MLDS (appearance method). The experiment included two separate experiments, where each one had the same 4 observers. Both experiments were tested with different parameters such as color, flicker frequency, and spatial frequency. Both experiments were conducted for three different spatial frequencies: 0.5 cpd, 1.0 cpd, and 8.0 cpd. The 0.5 cpd experiment was conducted with 8 Hz flicker, and the 1 cpd and 8 cpd experiments had no flicker.



(a) **Observers were asked to choose which presented stimulus had higher contrast, the right one or the left one.**



(b) **Observers were asked to choose, which of the bottom two stimuli had a higher contrast difference compared to the one on the top.**

Figure 7: **The two experiments that were conducted by Shooner and Mullen (2022) are shown in the figure. The first experiment (a) was a contrast discrimination experiment. The second experiment (b) was a contrast difference scaling experiment.**

2.4.1 Contrast Discrimination

In the contrast discrimination experiment, Shooner and Mullen (2022) investigated the relationship between pedestal contrast and increment thresholds. Pedestal contrast is defined as the physical representation of contrast in a stimulus. Increment thresholds are defined as the minimum needed contrast increment (x), where 75% of the time observers are can notice that the two stimuli differ. In other words Δx contrast needed for observers to notice the contrast difference between the two stimuli with contrast x and $(x + \Delta x)$. In the process of threshold measurement, observers were asked to choose the stimulus with the highest contrast (Figure 7a) and so performance was measured.

The result of the first experiment of an observer is shown in Figure 8. The relationship between pedestal contrast and increment thresholds is shown in log-log relation. The results replicate the findings of a “dipper function” as Bird et al. (2002). In Figure 8 the measured slope after the dip was 0.64. This shows that when the pedestal contrast is in the suprathreshold region the increment threshold is proportional to the pedestal contrast with a constant factor (the slope of the graph). According to the authors, this linear relationship can be found through different tested stimuli parameters. The slope value varied slightly but always existed according to the results of the experiment. The observed difference in the results was that the dip of each graph had a different location (threshold). The spatial frequency did not affect on the existence of this linearity as long as the pedestal contrast is in the suprathreshold region.

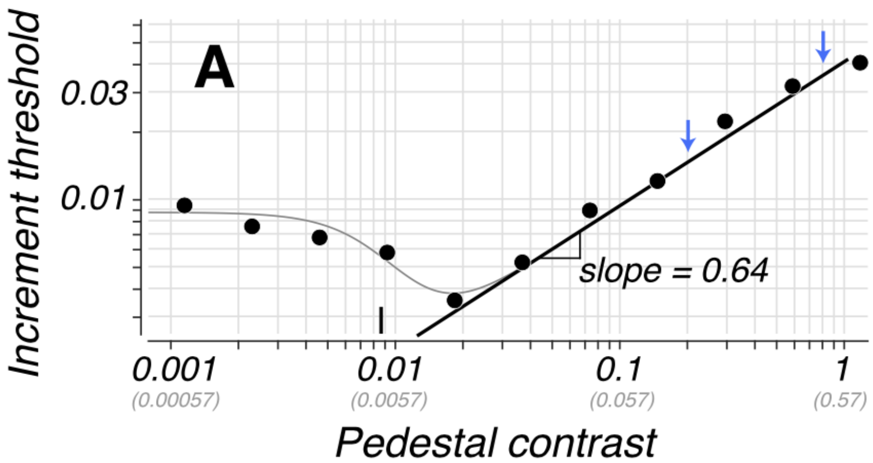
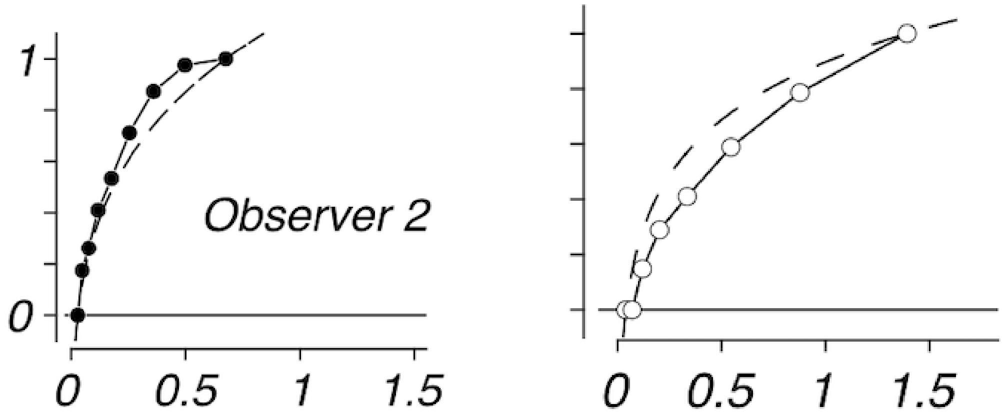


Figure 8: **The figure shows the result of a contrast discrimination experiment of an observer. The x-axis shows the pedestal contrast in log values and the y-axis shows increment thresholds in log values. Taken from Shooner and Mullen (2022).**

2.4.2 Contrast Difference Scaling Experiment

In the contrast difference scaling experiment (Figure 7b), the relationship between perceptual contrast and spatial frequency was investigated. The authors used the “Maximum Likelihood Difference Scaling” method which is an appearance measurement method to estimate perceptual scales. In the MLDS method observers are asked to choose the highest difference to an example stimuli and the number of options is at least two. In this experiment, observers were asked which stimulus was more different than the top middle one, the left bottom one, or the right bottom one (Figure 7b).

The result of the contrast difference scaling experiment for two spatial frequencies is shown in Figure 9 with dotted lines. Figure 9a shows the results for 1 cpd and Figure 9b shows the results for 8 cpd. Both shown results are from the same observer. The scales were normalized by the assignment of the minimum contrast to 0 and the highest contrast to 1. On the x-axis contrast and on the y-axis perceptual contrast is visualized. The dashed line shown in the figure will be explained below. The scale for the 1 cpd experiment was steeper than the experiment for the 8 cpd for this observer.



(a) **Tested parameter:1 cpd.**

(b) **Tested parameter: 8 cpd.**

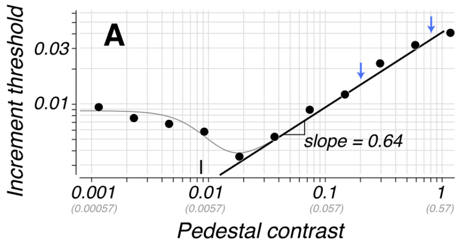
Figure 9: **In both figures, the relationship between perceived contrast and physical contrast is normalized and displayed in dotted lines. The cut line in the figure will be explained in chapter 2.4.3. The figures are taken from Shooner and Mullen (2022).**

2.4.3 Scale from the Contrast Discrimination Experiment

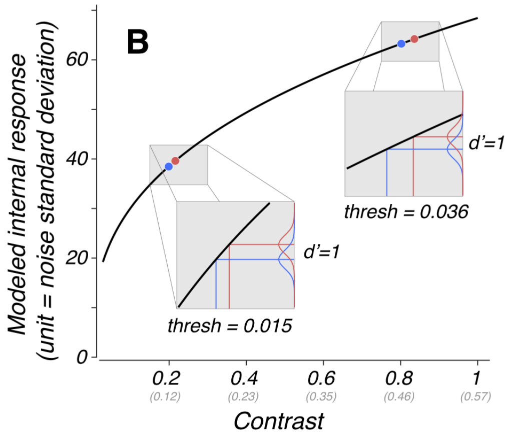
Shooner and Mullen (2022) created these two experiments to investigate a possible relationship between contrast discrimination (performance) and contrast difference scaling (appearance). The authors predicted the results of the contrast difference scaling experiment by using the slope of the linear part of the contrast discrimination experiment result. This is called **discrimination scaling** which is a procedure to derive scales from discrimination data, taking the assumption of additive noise (see section 3.3 for more details).

In Figure 10, it is shown how the named model is used by creating the scale on Figure

10b from the discrimination experiment result (the linear part) Figure 10a. The methods for creating the predicted scale will be explained in detail in the following section 3.3.



(a) **The relationship between pedestal contrast and increment thresholds is shown in 1 cpd spatial frequency in log-log relation.**



(b) **The relationship between physical contrast and perceptual contrast is shown.**

Figure 10: **Figure (a) shows the result of the contrast discrimination experiment (Figure 7a). Figure (b) shows the modeled predicted results of the contrast scaling experiment (Figure 7b). Taken from Shooner and Mullen experiment (2022).**

Shooner and Mullen (2022) furthermore applied this model to all the contrast discrimination results and predicted a scale for the discrimination scaling experiment. The Figure 11 shows a comparison between the actual and predicted scales for all tested parameters. Most of the cases show that the modeled results were similar to actual results (all observers for 1 c/deg 0 Hz in Figure 11). There were some cases where the predicted and the actual scale differed more (observers 2, 3, 4 for 8 c/deg 0 Hz in Figure 11). The agreement varied through different parameters and the lowest accuracy was 8 c/deg without flicker, which was not the case for 8 c/deg with 8 Hz flicker for all observers.

The complete result of the experiment is shown to emphasize that among all tested parameters the accuracy was lowest on the parameter spatial frequency (Figure 11). Precisely at 8 cpd 0 Hz the modeled perceptual contrast and the actual contrast difference scaling experiment result diverges.

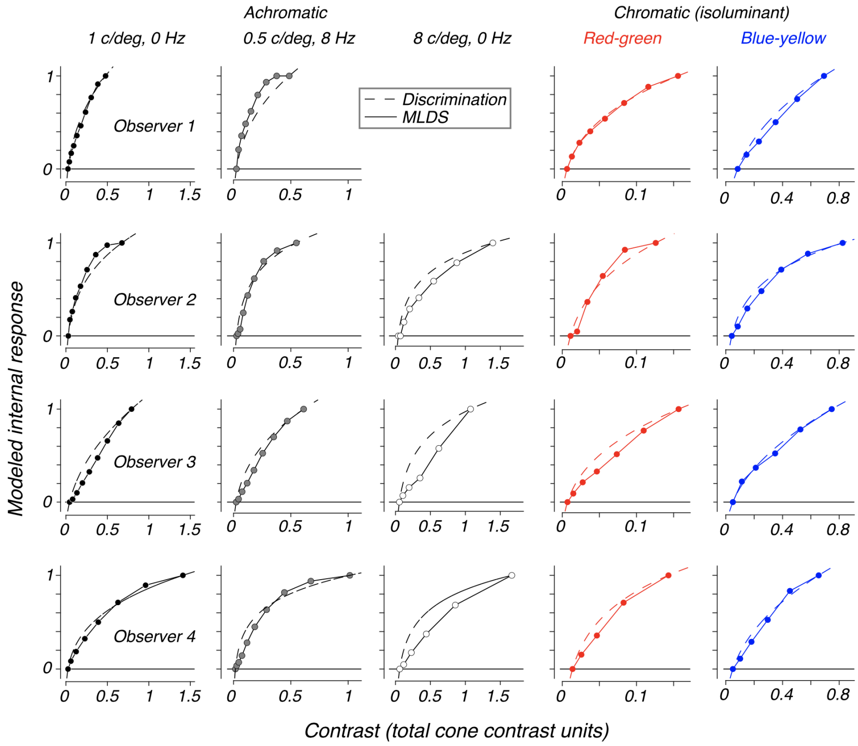


Figure 11: **The figure shows the result of a contrast discrimination scaling experiment for all tested parameters (dotted line) and the predicted scale created by the discrimination experiment (non-continuous line). Taken from Shooner and Mullen (2022).**

2.5 Research Question and Motivation

The relationship between discrimination and scaling experiments can be mentioned as a representative example of the difference between "Performance versus Appearance". In the domain of contrast, spatial frequency is researched through the years for discrimination and scaling experiments (Campbell and Robson (1968), Schade (1956), Georgeson and Sullivan (1975), Skottun et al. (1987), Bird et al. (2002)). All of these scientists found four major points regarding discrimination and scaling. Firstly, the visual system can discriminate low contrast stimuli better than high contrast stimuli (e.g.the threshold at contrast 0.8 is greater than the contrast at 0.1). Secondly, the contrast sensitivity varies in a parabola-shaped form in relation to spatial frequency (highest and lowest spatial frequencies have lower contrast sensitivity and therefore higher thresholds in comparison to mid-level spatial frequencies, where sensitivity is the highest). Thirdly, the difference between perceived contrast and physical contrast is larger at low and high frequencies than at mid frequencies. Lastly, the slopes of the "dipper function" in contrast discrimination experiments are similar or equal for different spatial frequencies.

To study the relationship between performance and appearance experiments, Shooner and Mullen predicted the results of an appearance experiment from the data of a performance experiment. The authors tested their model only with two spatial frequencies. The tested frequencies were 1 cpd and 8 cpd, at which sensitivity is not so different (CSF Figure 4) and perceived contrast is also similar (Contrast Constancy experiment Figure 5). In these two spatial frequencies the accuracy of the predicted scale varies heavily and both CSF and Contrast Constancy experiments showed that at a bigger spatial frequency difference (0-25 cpd) stronger effects can be observed.

It is still an open question whether the agreement found by Shooner and Mullen also apply to other spatial frequencies, taking into consideration that contrast constancy holds, scales should be different for different spatial frequencies.

This thesis proposes to use the discrimination scaling method on the results of the contrast discrimination experiment by Bird et al. (2002) to test how well discrimination data can be used to predict the appearance on the parameter spatial frequency. By testing 0, 2, 4, 8, 16 cpd (approximation of the used spatial frequencies in Bird et al. experiment), the different threshold and perceptual contrast levels from CSF and Contrast Constancy experiment will be investigated.

The present investigation of the discrimination scaling method will proceed as follows. First, the results of the contrast discrimination experiment (Bird et al. (2002)) will be digitized, the slopes of these digitized results will be measured and the results will be replotted. Second, scales will be created in two ways, using the slopes of the "dipper function" and using the raw data. Lastly, these scales will be compared to each other, to scales reported in the literature, and to the results of the contrast difference scaling experiment by Berlin Technical University Psychophysics lab in 2022.

3 Methods

3.1 Digitizing the Data

Bird et al. (2002) have published the results of the experiments as graphs. An external app called "Engauge Digitizer" was used to read out the data from a published figure. This software turned the figure into a CSV file, which was plotted as Figure 16a with four steps. The first step was importing the image of Figure 6 into the app. The second step was choosing the relationship between the variables (log-log). The third step was giving the three coordinates to show the app where the axes were. The last step was choosing the points and exporting the result as a CSV file.

3.2 Denormalizing the Data

Bird et al. (2002) reported values normalized by the observer's threshold. It was thus necessary to denormalize them. They have stated that the values for normalizing the results were the "best estimate of the contrast that corresponded to 75%" for each observer and were given in the original text. In the denormalizing process, these values were multiplied by each threshold and plotted again (Figure 16b). This denormalization was done for both observers separately (Figure 17b).

3.3 Discrimination Scaling

Creating a perceptual scale from a discrimination experiment is one of the earliest scaling methods developed, which goes back to Fechner (1860). The figure on the left (Figure 12, left one) shows a hypothetical contrast discrimination experiment result that fits Weber's law. On the x-axis the hypothetical physical variable S and on the y-axis the thresholds for this variable S are shown. Weber's law states that the thresholds (Δs) have a linear relationship with the physical variable s with a constant factor k as shown on the left of Figure 12. In this figure, the mentioned constant k is 0.1 meaning that the thresholds (Δs) are always equal to one-tenth of the physical variable (s). In other words, to reach each increase of i on thresholds (Δs) the physical variable s has to be increased $k*i$ (in this figure $0.1*i$). This constant increase of the thresholds (Δs) can be observed on the x-axis of the panel on the right (Figure 12).

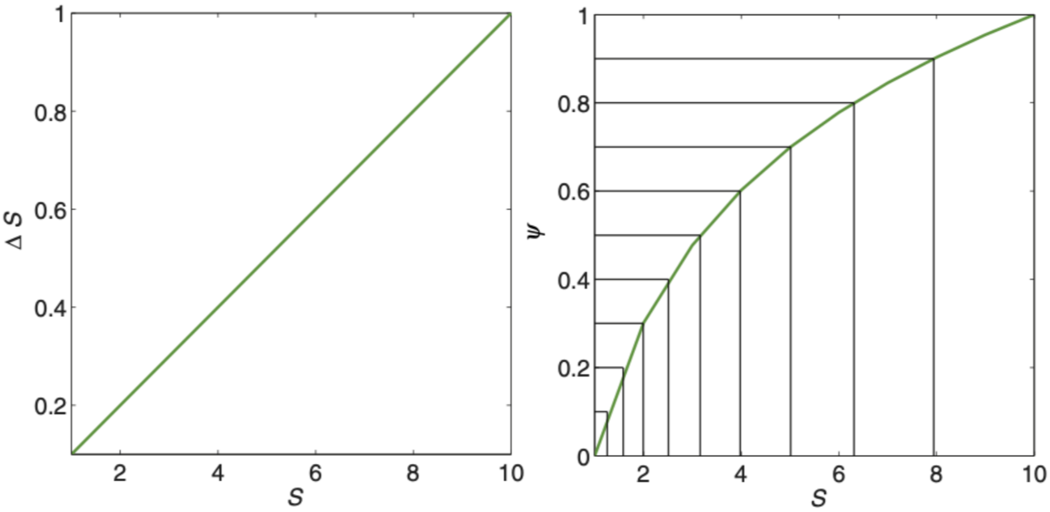


Figure 12: **The figure on the left shows a modeled discrimination experiment result created with Weber's law, where Δs is proportional to the hypothetical physical variable s. The figure on the right shows the scale created from the first figure using Fechner's law, where $\Psi(s)$ is proportional to the logarithm of the hypothetical physical variable s. Taken from Prins et al. (2016).**

To drive a perceptual scale, thresholds (Δs) are integrated along the stimulus variable (s). After the integration, the outcome is a perceptual scale that follows the natural logarithm ($\ln(s)$) of the physical variable s. This is known as "Fechner's law". Fechner's law states that perceptual points are proportional to the logarithm of the physical variable s with a constant k'. The right panel in Figure 12 shows the resulting Fechner's law derived scale from "Weber's law".

3.3.1 Mathematical derivation

The discrimination scaling method assumes that the constant increase in the thresholds exist in the perceptual response as well, meaning that the equal steps on the y-axis of the contrast discrimination result produces equal increments (new constant k') in perception ($\Delta\psi$). With the new constant k', the increments in perception ($\Delta\psi$) are equal to $k' * \frac{\Delta s}{s}$ and because the increment in thresholds and increment in perception are extremely small when their limit is considered, $\Delta\psi = k' * \frac{\Delta s}{s}$ turns into $d\psi = k' * \frac{1}{s} ds$. And the equation becomes $\psi = k' * ln(s) + C$, when both sides are integrated and c is a constant therefore can be considered as 0. This result is called "Fechner's law".

3.3.2 Numerical derivation

The mathematical derivation investigates a known function $(\frac{n}{x})$, in experiments, the obtained data does not necessarily follow a well-derived function. In Figure 13 different relationships between variables are shown. Although the mathematical derivation may be a good fit for data such as the bottom left panel it does not fit well with all other shown panels in Figure 13. The thresholds from the experiment are data points meaning the linear function used to summarize it may not be a good fit such as shown in the top right panel in Figure 13 and therefore the slope used to derive scales can be inaccurate. An analytical model is introduced to solve this problem. With this model, the scales are derived by interpolating each data point pair and therefore the need for summarizing the data with a linear function is eliminated.

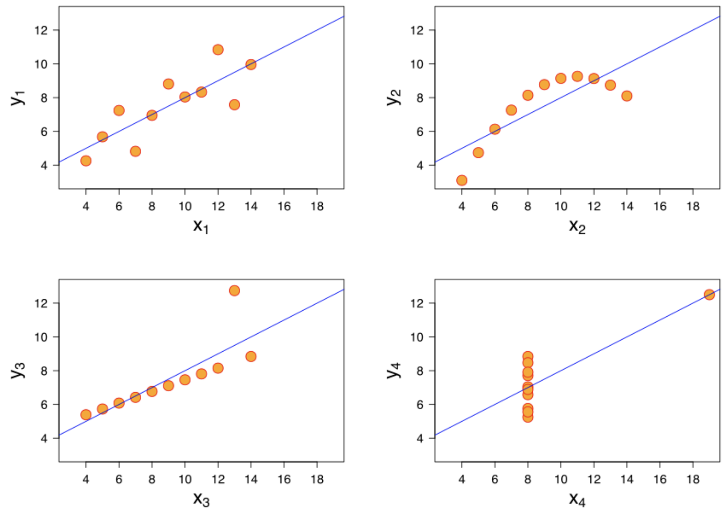


Figure 13: **The figure shows different relationships between two variables. Taken from Wikipedia [3].**

The algorithm of the numerical derivation is shown in Algorithm 1 as pseudocode. The Figure 14 shows a discrimination scale (Figure 14b) derived from the result of an hypothetical discrimination experiment (Figure 14a) by applying Algorithm 1. Discrimination scales are derived by doing two steps (the steps before the while loop in Algorithm 1). The first step is assigning the detection threshold (Δx_0 in Figure 14a) to the x-axis of the discrimination scale (x_0 in Figure 14b) and assigning 0 to the y-axis of the discrimination scale. The second step is calculating the next x-axis value of the discrimination scale (Δx_2 in Figure 14b) and incrementing the y-axis of the discrimination scale by one (or any constant value). The next x-axis value of the discrimination scale is calculated

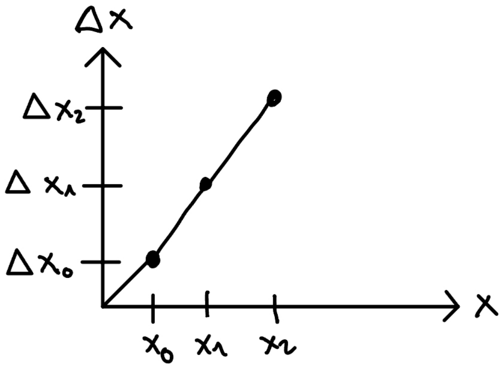
[3]https://en.wikipedia.org/wiki/The$_{d}ress/media/File : The_{d}ress_{b}lueblackwhitegold.jpg$

by adding the interpolation of the contrast discrimination result with the increment of the last added x-axis value (first threshold in the first iteration) to the last added x-axis value. This readout is iterated until the last added point to the x-axis of the discrimination scale (Figure 14b) is larger than the maximum value of the x-axis in discrimination result(Figure 14a).

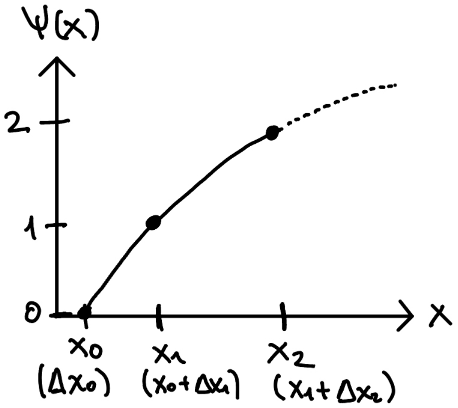
To sum up, each step done in numerical derivation (while loop in Algorithm 1) is the perceived variable in Figure 14b) and the x-axis of the Figure 14b) is the physical variable values calculated with the Algorithm 1 by the thresholds from Figure 14a).

Algorithm 1 Numerical derivation

```
data ← Data                                   ▷ Assign the data
scaleX ← [x_0]                                ▷ Add the detection threshold to x scale
scaleY ← [0]                                  ▷ Add 0 to y scale
contuniuing ← True
while contuniuing == True do
    c ← scaleX[−1]                            ▷ assign c the latest added element to x scale
    if c ≤ max(dataX) then                    ▷ if c is smaller than the max of x-axis of the data
        ΔX ← interpolate(c, dataX, dataY)     ▷ interpolation of data with c steps
        scaleX.append(ΔX + scaleX[−1])
        scaleY.append(1 + scaleY[−1])
    else
        contuniuing ← False
    end if
end while
```



(a) **Discrimination data.**



(b) **Perceptual scale.**

Figure 14: **(a) shows a hypothetical discrimination experiment result. On the x-axis the physical variable x is shown and on the y-axis the thresholds Δx are shown. (b) shows a discrimination scale derived from the result of a hypothetical discrimination experiment result (a). The shown hypothetical discrimination experiment result is linear for simplification purposes.**

4 Results

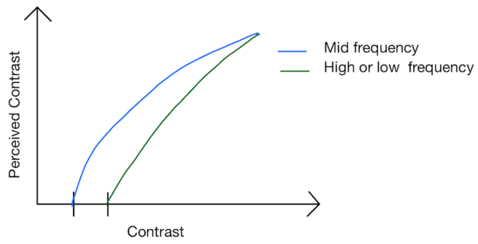
First, predictions about the scale pattern will be made according to different models mentioned in the literature. Second, the digitized data from Bird et al. (2002) will be plotted plainly and normalized by detection thresholds. Third, discrimination scales created from the discrimination experiment (Bird et al. (2002)) will be plotted. These scales differ firstly on whether the whole data or only the part that has a linear relationship is used.

4.1 Results Prediction

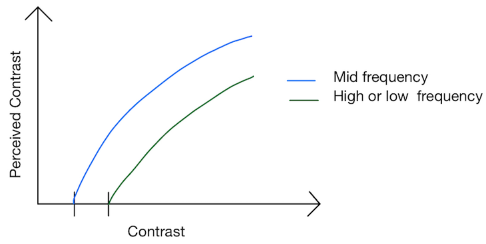
For simplicity, the stimuli of the contrast discrimination experiments will be classified according to their spatial frequency. The expectations of the results are categorized into “mid” and “high and low” spatial frequencies. The reason for this categorization is that the high and low spatial frequency have similar behavior in contrast discrimination (similar slopes after the dip) and scaling experiments (similar steepness). According to the slope of the contrast discrimination experiments, simulated scales are expected to have two types of behavior. Figure 15a and Figure 15b show two possible simulated results for a scaling experiment for the relationship between physical contrast and perceived contrast. In both figures, the mid-spatial frequencies are shown in blue, which have lower thresholds than high or low frequencies(shown in the green graph in Figure 15a and Figure 15b) as known from the CSF (neutral line on the x-axis).

If the contrast constancy holds, then with increasing physical contrast, perceptual contrast will be similar in all spatial frequencies as shown in Figure 15a. Since the thresholds from mid-level frequencies will be the lowest, and the perceptual contrast will be similar after a certain contrast value, scales for mid level frequencies will have a different shape, overall with a lower slope in comparison to high or low frequencies as in 15a.

On the other hand, if the only determinant of perceived contrast is contrast discrimination, and given that there is no difference across spatial frequencies (as Bird et al. (2002) and Skottun et al. (1987) have found), then the scales will look like the simulation in the Figure 15b. The thresholds from mid level frequencies will be lowest and scales will have similar slopes as in the Figure 15b.



(a) **Contrast constancy prediction.**



(b) **(Almost) Equal slopes prediction.**

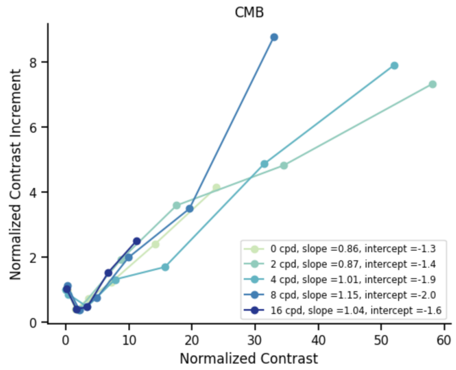
Figure 15: **The figures represent two different potential outcomes of the scales. The blue graph shows mid-spatial frequency and the green graph shows the high or low spatial frequency. (a) represent the expected scales, if contrast constancy holds. (b) represents the expected scales, if only contrast discrimination determines contrast appearance, given equal slopes in contrast discrimination.**

4.2 Data Replotting

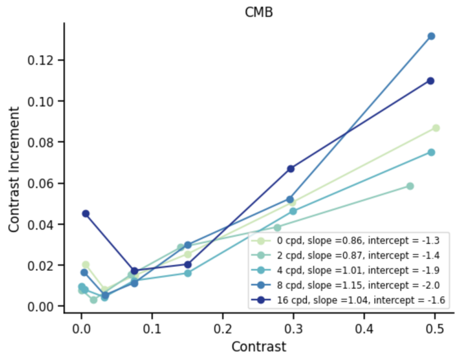
Figure 16 and Figure 17 show the plotting of the digitized results of the contrast discrimination experiment by Bird et al. (2002) as normalized and raw data. In the normalized and raw data results for both of the observers the so-called “dipper function” is to be spotted. For all spatial frequencies, the contrast increment decreases with increasing contrast until a minimum and then increases in an (approximately) linear way. For slope measurement (linear part of the data) and the intercept measurement, each spatial frequency result was fitted into a function. The measured slopes and the behavior of the raw data and normalized graphs for both observers were close to each other. The slope variation was higher for the observer CMB (highest difference: 0.29) and the highest measured slope was 1.15 for 8 cpd and the lowest measured slope was 0.86 for 0 cpd (16).

Some of the digitized slopes were different from the ones from the original experiment but they were mostly in confidence intervals. For the observer CMB the slope at 16 cpd was 0.01 out of the confidence interval (1.05-1.92) and for the observer GBH, all of the slopes were in the confidence intervals reported by Bird et al. (2002).

In the raw data results the lowest thresholds (the dip) vary between 0.1 and 0.001, the lowest minimum threshold was measured for 4 cpd, and the highest minimum threshold for 16 cpd. For observer CMB the intercept of the spatial frequencies had a higher interval than the observer GBH (0.7 and 0.5). The lowest measured intercept was -2.0 (8 cpd) and the highest measured intercept was -1.3 (0 cpd).



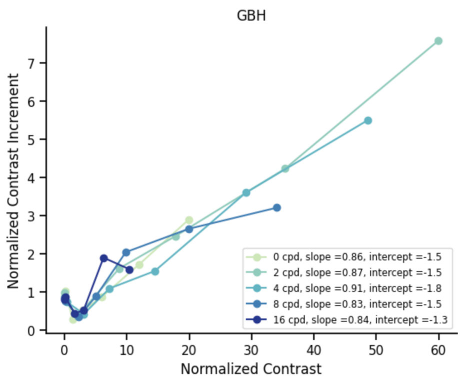
(a) **CMB normalized by the detection threshold.**



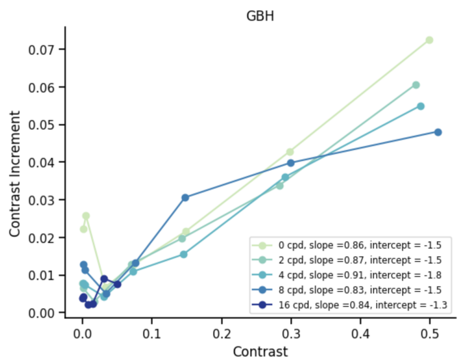
(b) **CMB raw contrast values**

Figure 16: **The results of the contrast discrimination experiment by Bird et al. digitized and plotted for the observer CMB. On the x-axis the normalized contrast is shown and on the y-axis normalized contrast increment is shown.**

The results for the observer GBH (Figure 16) showed that the measured slopes were very similar in each spatial frequency. The lowest slope was 0.83 for 8 cpd and the highest slope was 0.91 for 4 cpd. The lowest measured threshold interval for all spatial frequencies was between 0.001 and 0.01. The maximum value of the lowest threshold was measured for 16 cpd and the minimum value of the highest threshold was for 0 cpd. For the observer GBH the highest intercept was measured -1.3 at 16 cpd and the lowest intercept was measured -1.8 at 4 cpd. For both of the observers, the detection threshold order was the same. The order from the lowest to highest was: 2 cpd, 4 cpd, 8 cpd, 0 cpd, and 16 cpd.



(a) **GBH normalized by the first threshold.**



(b) **GBH Raw Contrast Values.**

Figure 17: **Same as Figure 16 but for observer GBH.**

To summarize, results for both observers differed in points such as the highest and the lowest measured minimum threshold (dip of the “dipper function”), highest and lowest measured slope spatial frequency. The behavior of the graphs (being a “dipper function”) and the detection threshold order for all spatial frequencies were the only two common aspects.

4.3 Scales derived from the discrimination experiment by Bird et al.

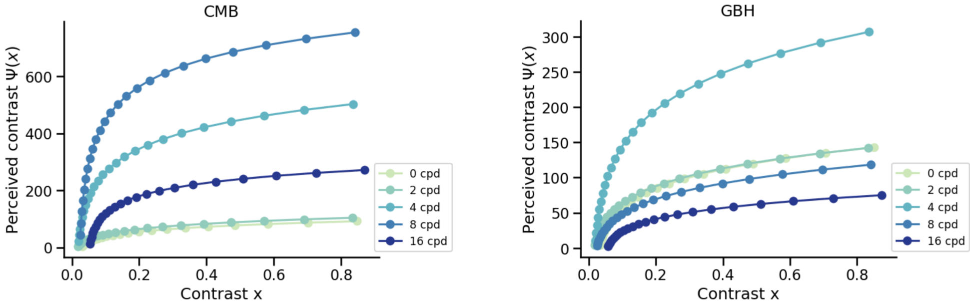
In this section previous digitized data from Bird et al. (2002) will be used to derive discrimination scales. All scales have contrast (x) in the x-axis (the first point is the first threshold in the discrimination data) and perceptual contrast $\Psi(x)$ on the y-axis and colors represent the spatial frequencies. The observer CMB is shown on the sub Figure (a) and the observer GBH is shown on the sub Figure (b) in all scales. Scales were derived in two ways: using only the linear part of the “dipper function” (similar to Shooner and Mullen (2022)), or using the whole function. In both cases, the numerical method described in section 3.3.2 was used.

4.3.1 Scales derived from the linear part of the “dipper function”

In this section, scales are derived by applying the discrimination scaling method to the linear part of the data and normalizing the scales in two different ways for each observer: normalization to the maximum value among all spatial frequencies and normalization for each spatial frequency to its maximum value.

First result 18 is achieved by applying the discrimination scaling method to the slope of the contrast discrimination experiment result of Bird et al. (2002) for both observers. The slope is calculated only with the data points with a linear relationship (right side of the minimum). The y-axis values are the steps taken during the discrimination scaling method, which represents perceived contrast in an arbitrary scale. As these values are equal $\Delta\Psi$, a scale with a higher maximum means that for conditions higher differences are perceived.

In Figure 18a the scale for 8 cpd is the steepest and the scale for 0 cpd is the shallowest with a high difference of perceived contrast at high contrast. In all spatial frequencies, perceived contrast difference increases rapidly after approximately 0.05 contrast and the high difference of perceived contrast is spotted in the highest contrast values as well. In Figure 18b the steepest scale is for 4 cpd and the most shallow scale is for 16 cpd. The highest perceived contrast at high contrast differs only for 4 cpd. The other spatial frequencies have similar perceived contrast at high contrast.



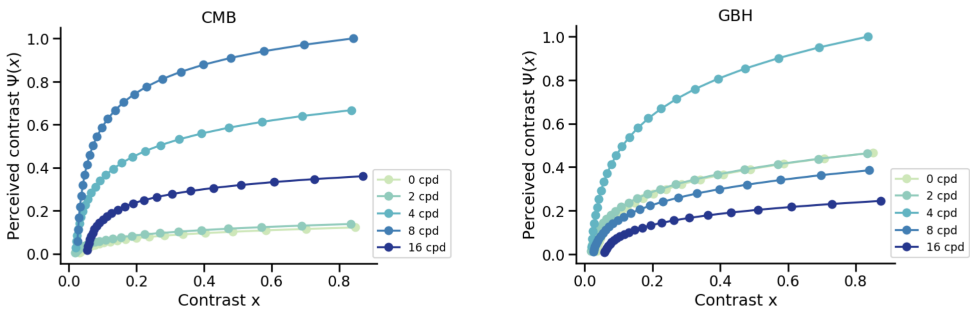
(a) **Scales created using discrimination data from observer CMB.**

(b) **Scales created using discrimination data from observer CMB.**

Figure 18: **Both figures are created by applying the discrimination scaling method to the result of the contrast discrimination experiment of Bird et al. (2002).**

The scales for the observers CMB and GBH differ on the steepest and the most shallow spatial frequencies. There is a difference in which spatial frequency is the highest perceived at high contrast (8 cpd for CMB and 4 cpd for GBH) and which spatial frequency is the lowest perceived at high contrast (0 cpd for CMB and 16 cpd for GBH) as well. Both results have the behavior that for each spatial frequency the perceived contrast differs among each frequency at low contrast and the difference is similarly large at high contrast as well.

A second visualization is created by normalizing the scales from 18 using the maximum value of perceived contrast across all spatial frequencies. This normalization does not change the pattern of the scales but changes the perceived contrast values. The goal of this normalization is to map a clear relationship between perceptual and physical contrast where perceived contrast now has a range between 0 and 1. The scale in the Figure 19 shows that for the scale of observer CMB at 8 cpd and for the scale of observer GBH at 4 cpd at high contrast perceived contrast is similar to physical contrast. But on mid-contrast perceived contrast is higher than contrast for both scales. On all other spatial frequencies for both scales, the perceived contrast is lower than the contrast after approximately 0.05 contrast.

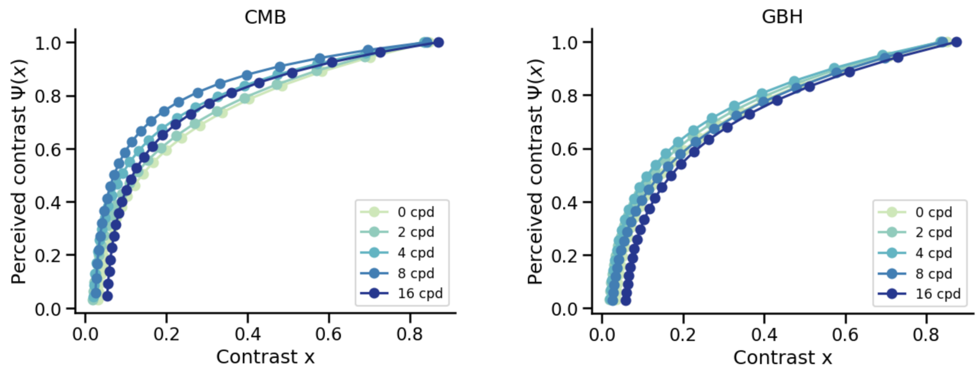


(a) **Observer CMB.** (b) **Observer GBH.**

Figure 19: **Same as Figure 18 but scales are normalized to the maximum, for each observer separated.**

A third result is created by normalizing the scales from Figure 18 by the maximum perceived contrast of each spatial frequency and therefore the maximum for all spatial frequencies are equal to 1. This normalization was done to analyze the shape of the scales, independent of their range.

The scales for the observer CMB are more different in shape from the scales for observer GBH. Figure 20a shows that the 4 cpd and 8 cpd spatial frequencies have steeper scales and 0 cpd and 2 cpd have more shallow scales than the other spatial frequencies. This is very different in Figure 20b, where all the spatial frequencies have almost equal steepness with 4 cpd being the steepest. Furthermore the scales in the Figure 20a are steeper than the ones in Figure 20b.



(a) **Observer CMB.** (b) **Observer GBH.**

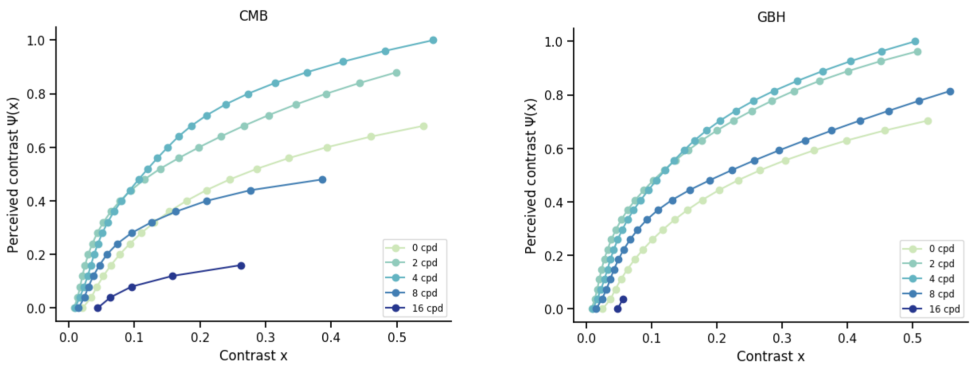
Figure 20: **Same as Figure 18 but scales were normalized to the maximum separately, for each observer and spatial frequency.**

4.3.2 Scales derived using the whole "dipper function"

In this section, the method used in the first three results is applied to the whole data and then normalized by the factors same as Figure 19 and Figure 20. The reason for creating these scales was to evaluate whether a different result is obtained.

Here the same discrimination scaling procedure (section 3.3.1) is applied now to the whole data and normalizing it to maximum steps among all spatial frequencies (same normalization factor as Figure 19) meaning the maximum for all spatial frequencies is equal to 1.

Figure 21a shows the scales for the observer CMB. The highest perceived contrast was measured at 4 cpd among the scales and 2 cpd followed a similar pattern with a slightly lower end-value as 4 cpd. The lowest perceived contrast was measured at 16 cpd. The spatial frequencies 0 cpd and 8 cpd had mid-perceived contrast values but 0 cpd had a slightly higher end-value. Figure 21b shows the scales for the observer GBH. All scale patterns were similar to the observer CMB. The only difference was that the perceived contrast was higher at 8 cpd than 0 cpd for the observer GBH and this was opposite the case for the observer CMB. It is important here to note that the discrimination scaling allowed two points for 16 cpd for the observer GBH and therefore the behavior for this scale can not be analyzed.



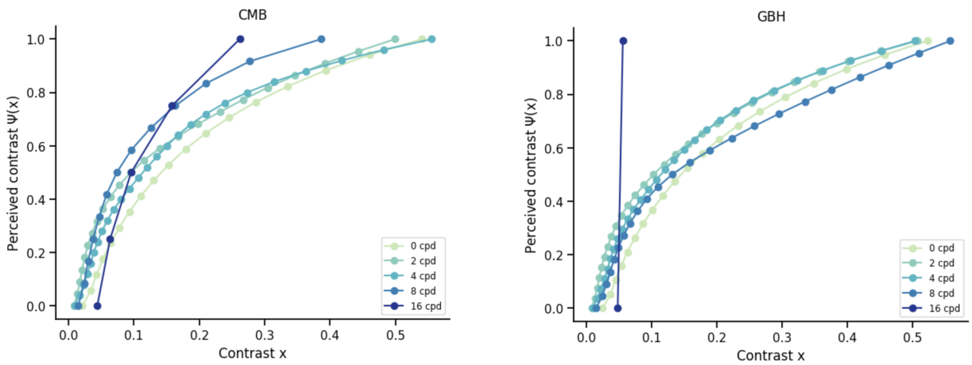
(a) **Observer CMB.** (b) **Observer GBH.**

Figure 21: **Both figures are created by applying the discrimination scaling method to the whole discrimination data of Bird et al. (2002) for both observers (not just the linear data points) and normalizing these points to the maximum steps required among all spatial frequencies.**

Scales for both observers have the same spatial frequency for lowest (16 cpd) and highest (4 cpd) perceived contrast. The pattern of the scales was similar for each spatial frequency as well. For both of the observers, the scales for 4 cpd and 2 cpd were similar in terms of the highest perceived contrast being close to each other and the pattern of the scales. The scales for frequencies 0 cpd and 8 cpd were similar in those two aspects as well but for observer GBH the scale for 8 cpd was significantly steeper than the scale for CMB. To conclude all the spatial frequencies had similar scale patterns for both of the observers with the except for observer CMB having a shallow scale for 8 cpd.

A fifth result was created by applying discrimination scaling to the whole data and normalizing each scale for the maximum steps (highest perceived contrast) of the spatial frequency (same normalization factor as Figure 20) meaning the maximum steps of each spatial frequency was equal to 1.

The scales for observer CMB had similar patterns for 0 cpd, 2 cpd, and 4 cpd (the shallowest three spatial frequencies, 0 cpd being the shallowest). The pattern for 4 cpd and 2 cpd were nearly identical. The steepest two spatial frequencies were 8 cpd and 16 cpd. For observer GBH all scales except 16 cpd had a similar pattern. The 16 cpd scale was the steepest. The shallowest scale was for 8 cpd.



(a) **Observer CMB.** (b) **Observer GBH.**

Figure 22: **Both figures are created by applying the discrimination scaling method to the whole discrimination data of Bird et al. (2002) for both observers (not just the linear data points) and normalizing these points to the maximum steps required for each spatial frequency.**

To sum up, the scales for 0 cpd, 2 cpd, and 4 cpd had the almost identical shape for the observer CMB and the scales 0 cpd, 2 cpd, 4 cpd, and 8 cpd had almost identical shapes for observer GBH. The shape for the 16 cpd was different from all other scales for both observers.

5 Discussion

In this section first, the scales from section 4.3 will be evaluated against the Result Prediction section (4.1). Second, the difference between the scales from the whole part of the data and the linear part of the data will be discussed. Third, the scales from section 4 will be compared to experiments from the literature, where the relationship between physical and perceived contrast is investigated. Lastly, the limitations of the procedure will be inquired, which will mention, how digitizing accuracy and the reliability of the results of the contrast discrimination experiment by Bird et al. (2002) can affect the results.

5.1 Results Comparison

5.1.1 The slope model of discrimination scales

In section 4.1 two different predictions of the results were described. The first prediction (Figure 15a) stated that the scales of high spatial and low spatial frequency must be lower than mid-spatial frequencies because the detection threshold is higher and at high contrast perceived contrast must be equal for all spatial frequencies (contrast constancy Georgeson and Sullivan (1975)). This result contradicts the findings of Bird et al. (2002) and Skottun et al. (1987), which state the correlation between the spatial frequencies and the slopes of the scales is close to zero. But this prediction relies on the slopes from mid-spatial frequencies being lower than the high or low spatial frequencies.

The second prediction (Figure 15b) stated that the scales from all spatial frequencies had equal slopes and therefore the mid-frequency levels would have the highest perceived contrast at high contrast because of the detection threshold being the lowest at mid or high-frequency levels would have a lower perceived contrast at high contrast because of the detection threshold being higher. This prediction contradicts the results of the experiment Georgeson and Sullivan (1975) (Figure 5), where at high contrast, the perceived contrast must be equal among all spatial frequency scales.

Next, the scales created from different parts of the data normalized to the maximum of the data will be discussed. The raw scales will not be compared to result predictions, since the discrimination scales have arbitrary perceived contrast values the comparison would be the same as scales normalized to the maximum of the data. The normalization factor of the maximum of each spatial frequency is not optimal for the result because this normalization focuses on the shape and therefore is not relevant for result predictions.

Discrimination scales from linear part of the data

The scales created from the linear part of the data for all spatial frequencies (Figure 19) show that the slopes of the scales differ significantly for observer CMB. The scales for 0 cpd and 2 cpd have almost identical patterns and have the equal perceived contrast at high contrast as Contrast Constancy states. At high contrast perceived contrast is not equal for any of the other scales and therefore it is a contraindication for Contrast Constancy. These scales do not fit the Equal Slopes Model shown in Figure 15b as well. As opposed to the Equal Slopes Model firstly the detection thresholds for all these scales are extremely close to each other. Secondly, the slopes are significantly different from each other meaning that at high contrast the spatial frequency with the lowest detection threshold (2 cpd) was not the highest perceived contrast.

For observer GBH the scales (Figure 19) show a combination of both of the models shown in Figure 15. The scales for 0 cpd and 2 cpd are almost identical patterns and have the equal perceived contrast at high contrast and therefore fit into both models. The perceived contrast at high contrast for 8 cpd and 16 cpd are close to 0 cpd and 2 cpd. These four scales have nearly the same perceived contrast at high contrast and therefore fit into the model shown in Figure 15a (Contrast Constancy Model). The fifth scale for 4 cpd has significantly higher perceived contrast than the other four scales and therefore the five scales fit more into the model shown in Figure 15b (Equal Slopes Model). An inverse correlation between the detection thresholds and perceived contrast at high contrast is reported during the investigation (example: 16 cpd lowest detection threshold lowest perceived contrast at high contrast).

For both of the observers, 0 cpd and 2 cpd have nearly identical patterns and therefore they have the same perceived contrast at high contrast, which fits into the Contrast Constancy Model and they have an equal slope, which fits into the Equal Slopes Model as well. For the observer CMB the scales do not fit into the mentioned two models. For the observer GBH the scales fit more into the Equal Slopes Model because the slopes do not correlate with the detection thresholds.

Discrimination scales from whole part of the data

The scales created from the whole part of the data for all spatial frequencies (Figure 21) for both observers show that at high contrast the scales have a large difference in perceived contrast and there is no correlation between the slopes and detection thresholds, which counters the Contrast Constancy Model (Figure 15a). Perceived contrast at high contrast is higher for mid-spatial frequencies and therefore the scales fit into the Equal Slopes Model (Figure 15b). The slopes vary for both observers by having higher slopes at 2 cpd and 4 cpd than the rest of the scales and therefore it is not a perfect fit for the Equal Slopes Model.

For all the scales (Figure 21) an inverse correlation between detection thresholds and perceived contrast at high contrast is spotted except for the spatial frequencies that have similar thresholds (for example 2 cpd has a slightly lower detection threshold than 4 cpd but perceived contrast is slightly higher at high contrast). Because of this inverse correlation and that perceived contrast is different at high contrast for each scale, the scales fit more into the Equal Slopes Model, despite having a small difference in the slopes of the scales. The scales for 2 cpd and 4 cpd have similar patterns and the scales 0 cpd and 8 cpd have similar patterns compared to each other as well. The scale for 16 cpd makes the mentioned inverse correlation easier to spot because the scale has a significantly higher detection threshold and perceived contrast is lower at high contrast than other scales.

An alternative model is proposed after investigating the scales (Figure 21) for both observers, which states that there is a weaker relationship between slopes and the detection thresholds but rather a stronger relationship between perceived contrast at high contrast and detection thresholds (inverse correlation). The detection thresholds are similar for 2 cpd and 4 cpd for both observers and the perceived contrast at high contrast is also similar. The same similarity is spotted for both observers between 0 cpd and 8 cpd where detection thresholds are similar. These spatial frequency pairs also have very similar slopes in comparison to each other but there is a large difference in comparison to other scales.

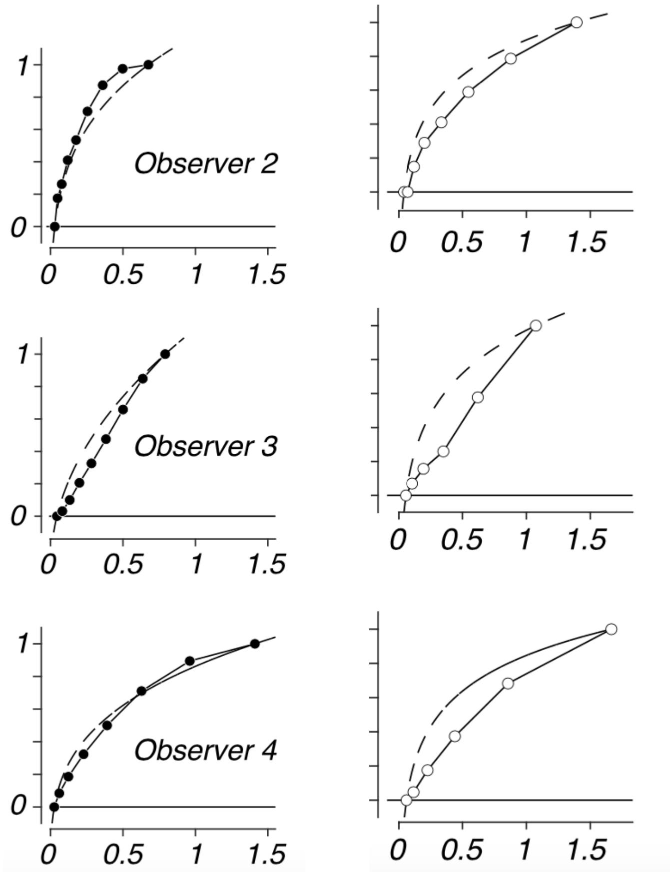
Comparison between the scales created from the linear and the whole part of the data

The relationship between detection thresholds and slopes of the discrimination scales were investigated in the Result Prediction section (4.1). The scales created from the linear part of the data (Figure 19) and the whole data (Figure 21) both show opposing behavior against Contrast Constancy Model (Figure 15a) because at high contrast perceived contrast among the scales has a large difference. On the other hand scales from both parts of the data agree with the Equal Slopes Model (Figure 15b). The scales created from the linear part of the data have a higher difference in slopes than the scales created from the whole part of the data, which questions their fit to the Equal Slopes Model. The scales created from the whole part of the data have similar slopes and the perceived contrast at high contrast is different among the scales and therefore is a good fit for the equal slopes model. During the investigation, it is discovered that the detection thresholds and the perceived contrast at high contrast have an inverse correlation among both methods and all scales.

5.1.2 Discrimination scales compared to scales from literature

This section includes three comparisons. Firstly, the scales from Shooner and Mullen are compared to each other, and to the scales shown in section 4.3. Secondly, the scales from Berlin Technical University Psychophysics Lab are compared to each other and to the scales shown in section 4.3.

In 2022 Shooner and Mullen designed a contrast discrimination and a contrast difference scaling experiment for two spatial frequencies: 1 cpd and 8 cpd. They have created discrimination scales from the linear part of the contrast discrimination experiment result as well and compared them to the results of their contrast difference scaling experiment. Figure 23 shows the discrimination scales shown in the non-continuous line and the contrast difference scaling experiment results (dotted line) for three observers. The original paper includes four observers but the first observer is not shown because the results do not exist for this observer for 8 cpd. For all of the observers, the discrimination scales at 1 cpd are more similar to the scaling experiment result than at 8 cpd meaning the discrimination scales are a better representation of the relationship between perceived and physical contrast at 1 cpd. The discrimination scales are steeper for observer 2 and observer 3 at 1 cpd than 8 cpd and for observer 4 they are similar to each other. The observers differed first in the steepness of the scales. The second difference in the scales was whether the discrimination scale has a higher or lower perceived contrast at mid-contrast than the scaling experiment results. The discrimination scales at 8 cpd being not similar to actual scales and discrimination scales at 1 cpd being similar to actual scales is the only common aspect for all observers.



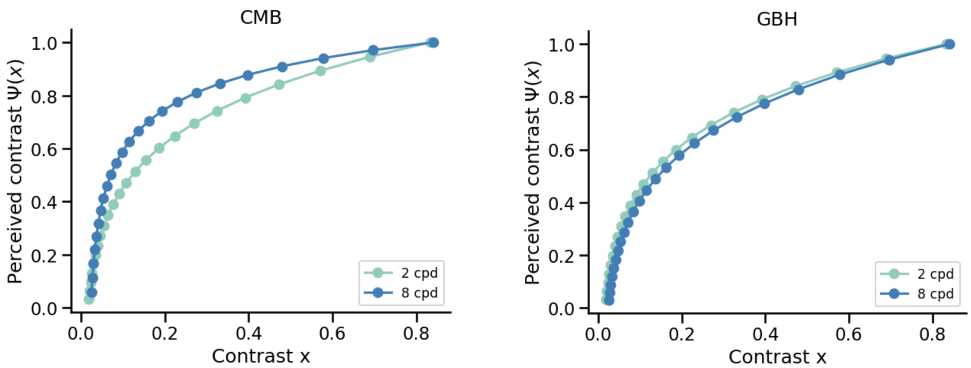
(a) **Tested parameter 1 cpd.**

(b) **Tested parameter 8 cpd.**

Figure 23: **Same as Figure 11 but only the panels focused on spatial frequency. Taken and modified from Shooner and Mullen (2022)).**

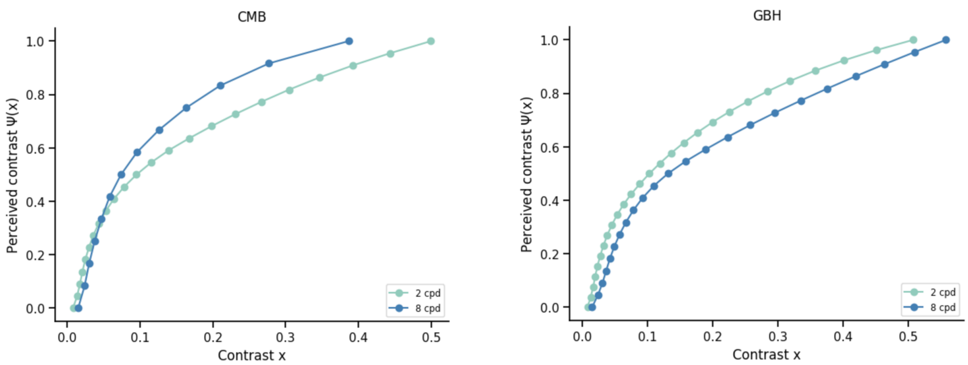
The contrast discrimination experiment results which were used to create discrimination scales in section 4.3 included 2 cpd and 8 cpd that are close to the spatial frequencies from Shooner and Mullen (2022) experiment by assuming that the difference of perceived contrast between 1 cpd and 2 cpd is low at all contrast levels (based on Contrast Constancy result shown in Figure 5). The discrimination scales from Shooner and Mullen (2022) (Figure 23) have a large difference in shape between the two spatial frequencies in comparison to the scales created from different parts of the data (Figure 24 and Figure 25). For observer GBH, the scales from the linear part of the data for both spatial frequencies are equal in shape. The same similarity exists for the scales created from the whole data (Figure 25) as well. Another difference between the scales from

Shooner and Mullen (2022) and section 4.3 is the variability among the observers. The shape of the discrimination scales for both spatial frequencies for observer 4 in comparison to the other two observers in Shooner and Mullen (2022). This difference did not exist in the scales for the observer CMB and GBH in the results for section 4.3. Furthermore, whether the whole data or linear part of the data was used during creating the scales had no effect during the comparison with the results from Shooner and Mullen (2022).



(a) **Observer CMB.** (b) **Observer GBH.**

Figure 24: **Same as Figure 20 but only for the spatial frequencies 2 cpd and 8 cpd (same as Shooner and Mullen (2022) experiments).**

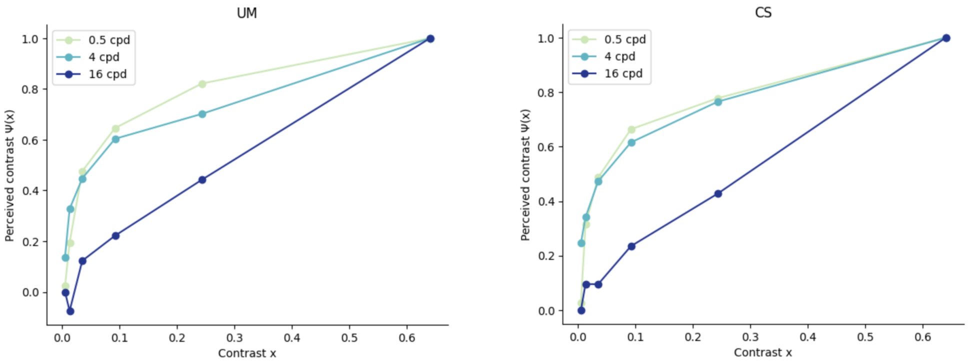


(a) **Observer CMB.** (b) **Observer GBH.**

Figure 25: **Same as Figure 22 but only for the spatial frequencies 2 cpd and 8 cpd (same as Shooner and Mullen (2022) experiments).**

Comparison to Berlin Technical University Psychophysics Lab results

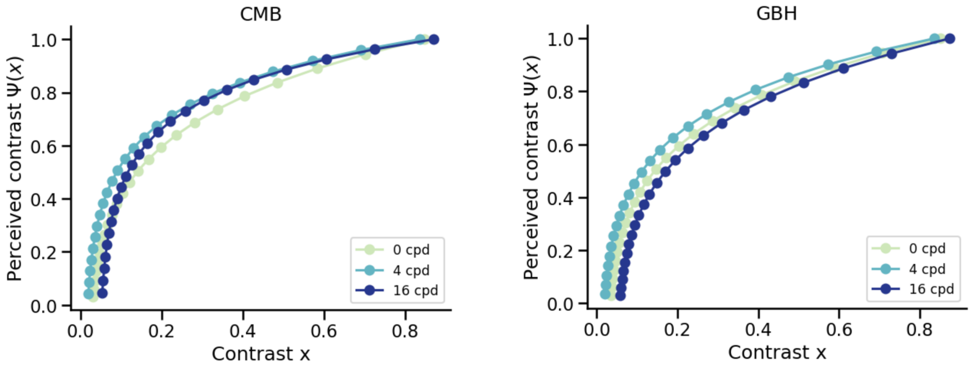
In 2022 Berlin Technical University Psychophysics Lab 2022 conducted a contrast perception experiment by using the “Maximum Likelihood Conjoint Measurement” method (an appearance method). Figure 26 (normalized to the maximum perceived contrast among all spatial frequencies same as Figure 19) shows the results of the experiment for two observers and the spatial frequencies 0.5 cpd, 4 cpd, and 16 cpd. The scales for 0.5 cpd and 4 cpd have a similar shape for both observers being the steepest two and 16 cpd differs from these two frequencies being the shallowest for both observers. There is not any significant difference for both of the observers (same spatial frequency as the steepest and shallowest shape).



(a) **Observer UM.** (b) **Observer CS.**

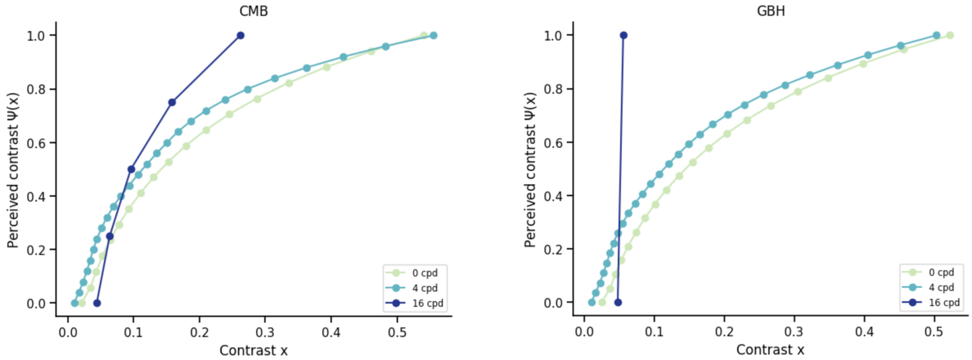
Figure 26: **The figure is created from Berlin Technical University Lab data by the same normalization factor as Figure 22. This data was measured using MLCM on two different observers (UM and CS).**

The scales created in section 4.3 and the scales from the experiment done by Berlin Technical University Psychophysics Lab have the following spatial frequencies in common: 0 cpd (assumed perceived contrast is similar with 0.5 cpd at all contrast levels), 4 cpd and 16 cpd. Figures 27 (same as Figure 19) and 28 (same as Figure 22) show the scales from section 4.3 (only the spatial frequencies 0 cpd, 4cpd, 16 cpd.) In the Figure 27 the scales differ in terms of shape for both observers which is different from the results from the Lab where for both observers the steepest (0 cpd) and the shallowest (16 cpd) spatial frequencies are equal. The discrimination scales for 0 cpd and 4 cpd have similar shapes as in the results from Lab but the scale for 16 cpd is extremely shallow compared to the discrimination scales for both observers.



(a) **Observer CMB.** (b) **Observer GBH.**

Figure 27: **Same as Figure 20 but only for the spatial frequencies 0 cpd, 4 cpd, and 16 cpd (same spatial frequencies as 26).**



(a) **Observer CMB.** (b) **Observer GBH.**

Figure 28: as Figure 22 but only for the spatial frequencies 0 cpd, 4 cpd and 16 cpd (same spatial frequencies as 26).

5.2 Limitations

In this section, the limitations that affect the results will be discussed. The first limitation was that the experiment results compared in section 4.1 were done with different experimental methods and different spatial frequencies. The second limitation encountered was whether the digitizing of the data may affect the results. The third limitation was whether the agreement between the thresholds and spatial frequencies in the Bird et al. (2002) was significant.

5.2.1 Difference in compared experiment results

The difference in spatial frequency in all the reviewed data is a possible limitation. Bird et al. (2002) have designed a contrast discrimination experiment for the spatial frequencies 0 cpd, 2 cpd, 4 cpd, 8 cpd, and 16 cpd, and the discrimination scales were created from the results of this experiment. Shooner and Mullen (2022) derived discrimination scales and contrast difference scaling experiments for 1 cpd and 8 cpd with the MLDS method. The only spatial frequency that exists in the Bird et al. (2002) experiment was 8 cpd (1 cpd is only assumed to be the same as 2 cpd) and these results are for different observers which weakens the comparison between the two experiments. Another aspect that weakens the comparison is that the scales from the Shooner and Mullen (2022) experiment differ significantly from each other and are only done for shape comparison (the maximum for each scale is equal to 1).

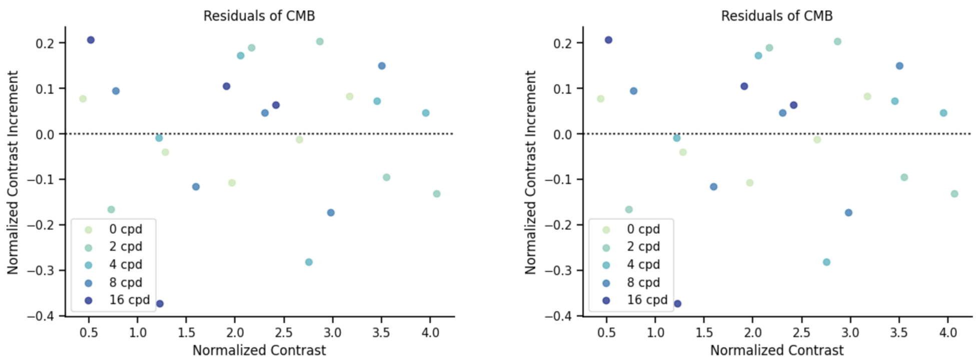
The experiments from Berlin Technical University Psychophysics Lab were done for spatial frequencies 0.5 cpd, 4 cpd, and 16 cpd with the MLCM method. All three scales existed in the Bird et al. (2002) experiment as well (0 cpd is assumed to be the same as 0.5 cpd). The Lab results are very consistent for the observers Um and CS in comparison to each other but the results only allow a shape analysis and it does not allow to compare them to the models mentioned in section 4.1. The shape comparison showed that only low and mid spatial frequencies agree with the discrimination scales in section 4.3.

5.2.2 Digitizing Accuracy

The results of Bird et al. (2002) included the figure results of the experiment but not the actual data. Therefore the actual data were digitized by using an external service. Many trials were done and witnessed that the accuracy of the digitized data was questionable but all the slopes were in the confidence intervals mentioned by the authors. For the observer CMB all the slopes of the digitized data were very similar to the original data for 0 cpd, 4 cpd, and 8 cpd. For 2 cpd the slope was slightly higher (0.2) and for 16 cpd it was significantly lower (0.5) than the original data. For the observer GBH, only slope of the digitized data that differed significantly (0.4 below) from the original experiment was 16 cpd. The reliability of the used external service was lower for log-log relations, which was used in the digitizing process.

The analysis done in section 5.1 relies on the reliability of the digitizing. This thesis assumes that the digitized data varies from the original experiment data but not significantly (shown by manually measured differences) except for 16 cpd for both observers. The difference between the digitized data and the original data was nearly in size of the confidence intervals reported by the authors. It is proposed that the analysis must be done with the actual data to trust the analysis 100%.

As a sanity check for the digitization process, I also evaluated the residuals of the linear fit done over the digitized points. Figure 29 shows these results. Cook and Weisberg (1982) mention that the residuals define the level of the agreement two data types can have. Residuals are commonly used for discrimination experiments in psychophysics to test if the measured variables have reliable relation to each other. Residual points being close to zero is the first aspect that shows a good agreement between the variables. The second aspect that shows a good agreement between the variables is that the residual points should have random points (no correlation). The spatial frequencies 0 cpd, 2 cpd, 4 cpd, and 8 cpd were reliable according to the two aspects mentioned. There was no correlation found for these spatial frequencies and the residuals points were around ± 0.2. This was not the case for 16 cpd. Even though the spatial frequency 16 cpd did not correlate for both of the observers, the residuals were not close to zero (± 0.4).



(a) **CMB** (b) **GBH**

Figure 29: **Residuals of the simulated contrast discrimination result.**

Bird et al. (2002) state that the results for both observers for 16 cpd were not reliable. This statement agrees with the results of the residuals investigation.

5.3 Open questions

The discrimination scales created from the results of Bird et al. (2002) experiment do not agree with the results of the Contrast Constancy experiment (Figure 5). At high contrast, the discrimination scales for all spatial frequencies do not have an equal perceived contrast as opposed to Contrast Constancy and the discrimination scales do not agree with contrast perception experiment results from the literature. The first possibility is that all three experiments mentioned in section 5.2 have different observers and a small number of observers. Furthermore, there is no experiment conducted at 25 cpd, where Contrast Constancy results show the equality of perceived contrast at high contrast among all spatial frequencies. The variability among the results of observers is high in the Shooner and Mullen (2022) experiment and therefore the result of the thesis can be affected. The second possibility is that Contrast Constancy does not hold and therefore the scales from different spatial frequencies will not have an equal contrast at high contrast. The last possibility is that discrimination scales can not predict the outcome of scaling experiments in the domain of contrast and performance and appearance experiments results are distinct from each other.

To investigate all the possibilities that affected the result of the thesis, a contrast discrimination and contrast perception experiments with the spatial frequencies from Bird et al. (2002) and Shooner and Mullen (2022) must be conducted with the same observers. For this experiment, a large number of observers would increase the reliability, since the results in the previous experiments for different observers have a large difference. In other words, an experiment conducted with spatial frequencies 0 cpd, 1 cpd, 4 cpd, 8 cpd, 16 cpd and 25 cpd are required to test whether performance experiments can predict the scaling experiments in the domain of contrast and to test whether Contrast Constancy holds. If the discrimination scales from this experiment have equal perceived contrast at high contrast then Contrast Constancy would hold. And if the contrast perception scales agree with the discrimination scales then performance experiments can predict the scaling experiments in the domain of contrast.

5.4 Conclusion

Discrimination scales are in contradiction with the prediction of contrast constancy. Second, the discrimination scales and the scales from the contrast perception experiment do not agree with each other, and using the whole part of the data or the linear part of the data for discrimination scaling makes no difference. This comparison shows that the behavior of the experiment scales and discrimination scales are different and therefore if contrast constancy is assumed to hold, contrast performance cannot predict contrast appearance. Lastly, the inverse correlation between detection thresholds and perceived contrast at high contrast was found through the investigation.

References

Aston, S., & Hurlbert, A. (2017). What# thedress reveals about the role of illumination priors in color perception and color constancy. *Journal of vision*, *17*(9), 4–4.

Bird, C., Henning, G., & Wichmann, F. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *JOSA A*, *19*(7), 1267–1273.

Bovik, A. C. (2009). *The essential guide to image processing*. Academic Press.

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of physiology*, *197*(3), 551.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.

Georgeson, M., & Sullivan, G. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of physiology*, *252*(3), 627–656.

Khaleefa, O., et al. (1999). Who is the founder of psychophysics and experimental psychology? *American Journal of Islam and Society*, *16*(2), 1–26.

Kukkonen, H., Rovamo, J., Tiippana, K., & Näsänen, R. (1993). Michelson contrast, rms contrast and energy of various spatial stimuli at threshold. *Vision research*, *33*(10), 1431–1436.

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology*, *25*(13), R545–R546.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Josa*, *70*(12), 1458–1471.

Prins, N., et al. (2016). *Psychophysics: a practical introduction*. Academic Press.

Schade, O. H. (1956). Optical and photoelectric analog of the eye. *JoSA*, *46*(9), 721–739.

Shooner, C., & Mullen, K. T. (2022). Linking perceived to physical contrast: Comparing results from discrimination and difference-scaling experiments. *Journal of Vision*, *22*(1), 13–13.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of neurophysiology*, *57*(3), 773–786.