IKV1 homework - Hypotheses Testing on ETRA Challenge dataset

Martin Picek

January 2022

1 Introduction

In this report we describe an experiment during which we measured a threshold where each subject perceives two patches to be of the same color. To be more specific, we observed a well known phenomenon - Simultaneous Contrast. In our experiment, we wanted to observe how much the contrast affects the perception and how consistent the subjects are in their choices during an experiment. This is why we also computed a confidence interval for the threshold.

1.1 Simultaneous Contrast

In our experiment, we work with an effect known for centuries - The Simultaneous Contrast - a phenomenon during which a color of a stimulus appears to shift towards the opposite color of its surroundings [Sae Kaneko et. al. 2018]. The first reference to this effect comes from the 11th century from Ibn al-Haytham [wikipedia.org] and it is under research since then.

Our experiment uses visual stimulus, but this phenomenon is not limited solely to this domain. John Locke, an English philosopher from the 17th century, observed that when a subject simultaneously places his right hand into hot water, his left hand into cold water, and afterwards he puts both hands into a container with lukewarm water, his right hand perceives the water as cold and the left hand perceives it as hot [Contrast in judgments of mental health, Kushner 2008]. Similarly, contrast plays a large role in our perception of art quality - whether we will like the music more or less depends on the music we heard right before [Music perception, Popper 2010].

Simultaneous Contrast is subconsciously perceived commonly, yet not identified as an usual phenomenon. Students are often frightened to be examined by a teacher when an excellent student was examined just before them. Their performance can be seen worse in comparison to the same performance but after a poor student's examination. In Czech, we also have a phrase that links to this - "Nasadit laťku" [in English: "set a bar/level"]. It is hard to be perceived as excellent if the quality is set too high prior to the act since the contrast degrades the perception of the performance.

1.2 The Experiment

The experiment consisted of two trials. Both trials were the same except from the color that was presented. In the first trial, the subjects (N = 3) were distinguishing which one of the two gray squares presented on the screen is brighter, in the second trial, the subjects were distinguishing between two red squares. The left square (the reference) had the luminance set to 0.5 and the right was set to a value ranging from 0.1 to 0.9 (in particular, values used were 0.1, 0.3, 0.4, 0.45, 0.5,

0.55, 0.6, 0.7, 0.9). There were 9 values to choose from in total. The background was solid black (in both experiments) and the left patch had a surrounding of the same color but of a lighter shade ($[1, 1, 1]$ for the gray trial and $[1, -1, -1]$ for the red trial). Program was presented full-screen on a standard 29-inch monitor with a fullHD resolution. The gray color had values $[x, x, x]$ and the red color had values $[x, -1, -1]$, where x (= the luminance) was set to one element of the previously mentioned set of values. Each value was repeated 30 times (30 times compared to the reference color of luminance set to 0.5).

Prior to the experiment, subjects trained with the testing environment for 1-2 minutes (under a supervision) to learn how to control the program. Right after that the first trial was conducted. A 10-20 minutes long break followed and after it the second trial with the red color was done.

The subjects had ids m, t and e. Subject m was a 21 year old man, t was a 56 year old man and the last subject e was 10 year old girl.

2 Results

For each subject we measured a threshold, where the subject is 50 percent likely to say that the right patch is brighter. Therefore this is the luminance that makes the right square appear the same as the reference square on the left.

As a psychometric function, we used $f(x) = \frac{1-\lambda}{1+e^{-\frac{(x-\alpha)}{\beta}}}$, where α, β and λ were the parameters that we learned using mean square error. In the function, the term $1 - \lambda$ stands for mistakes that a subject makes.

All the psychometric functions obtained are monotonic, therefore we could find the threshold with the following approach: We subtracted 0.5 from the function, so now the root of the function is the threshold. We found the root using the bisection method.

Similarly was found the 80% confidence interval. We were looking for such δ that holds $f(t + \delta) - f(t - \delta) = 0.8$, where f is our fit psychometric function and t is our already found threshold. From this we can obtain an equation: $f(t + \delta) - f(t - \delta) - 0.8 = 0$. Let's call the left hand side a function $g(\delta)$. We want its root. This function is monotonic and therefore we can use the bisection method to find its root, so that the equation holds. With this approach we obtained the 80% confidence interval.

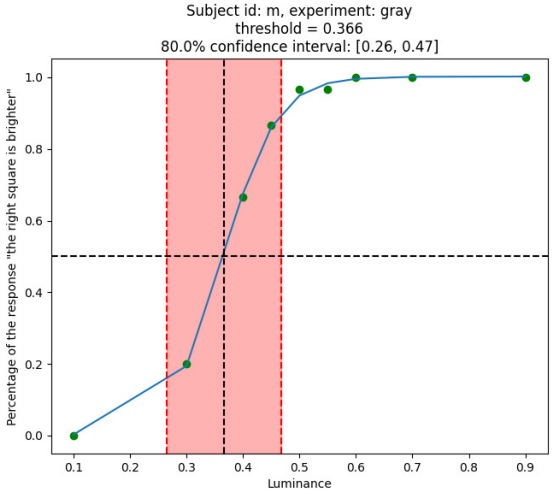
Here is a table with the results - thresholds and 80% confidence interval:

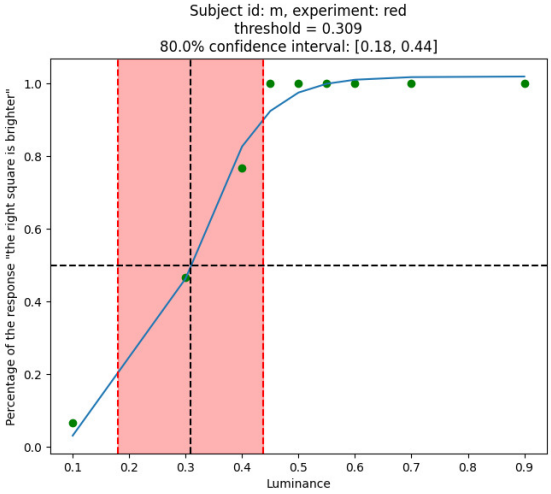
Trial	Subject	Threshold	80% interval
gray	m	0.366	[0.26, 0.47]
gray	t	0.405	[0.29, 0.52]
gray	e	0.348	[0.08, 0.62]
red	m	0.309	[0.18, 0.44]
red	t	0.406	[0.32, 0.49]
red	e	0.422	[0.21, 0.63]

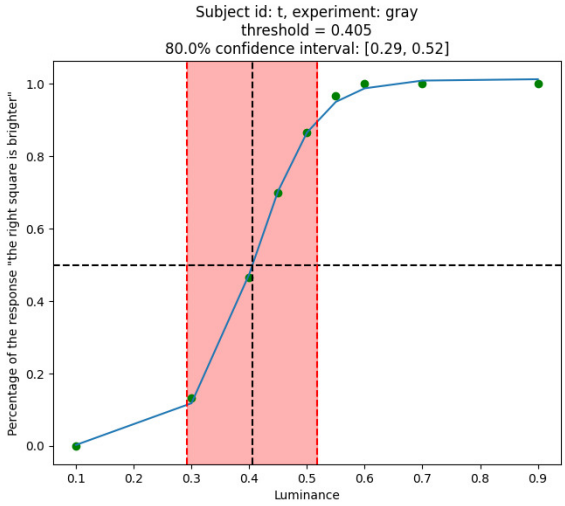
We can observe, that Simultaneous contrast really plays role. The average threshold for the first

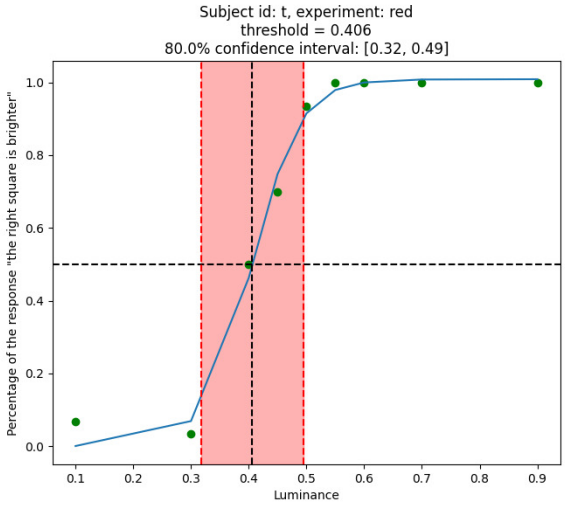
trial was 0.373 of luminance, for the second trial the average threshold was 0.379 of luminance.

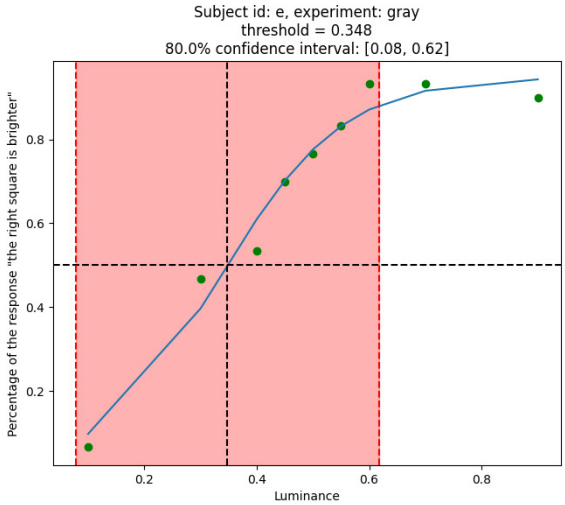
Subjects m and t were very consistent in their responses compared to the subject e. We were assuming that as the subject e was a 10 year old girl. We can observe it thanks to the width of the confidence interval.

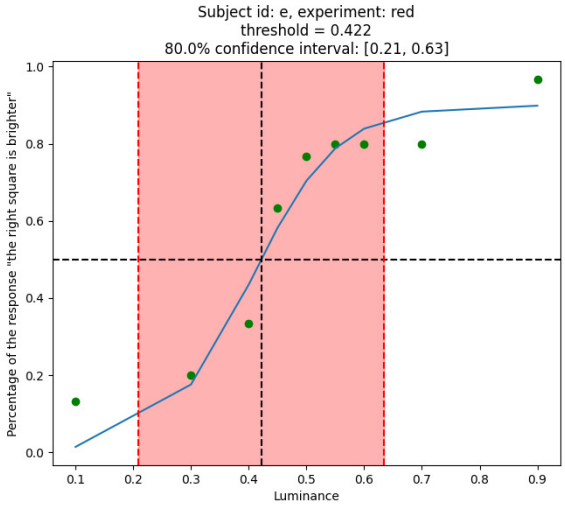












3 Conclusion and discussion

In this project, we tested the Simultaneous Contrast effect and how it affects perception of color of 3 subjects. We found out that the average luminance really deviates from the real one, therefore the effect played role significantly.

The subject e had a very wide confidence interval. It was a 10 year old girl, hence the possible explanation for that might by a lack of attention to the task. Other two subjects were considerably more consistent with their responses. No difference between the two trials was found.

For the future study, it might be better to work on the following points:

- Increase the number of subjects
- Increase a sample density of luminance around the presumed threshold. This study can be used as a prior for the assumption. We knew nothing about the threshold we were to obtain, thus we increased the density only around the point 0.5.
- If a subject has a very high variance, it might be better to choose a different subject. This is, however, a good topic for a discussion, since high variance in responses might also mean something interesting. If we omitted these subjects in our study, we might lose some important results.

It might be interesting to do the same experiment with different domains - for example the already mentioned perception of heat. In other domains, however, the experiment might be more difficult to conduct. It might be also interesting to somehow compare the Simultaneous Contrast in different domains - whether we have some sense that is more susceptible to be influenced by the observed phenomenon.