Will It Blend?

Studying Color Mixing Perception

Paulo Duarte Esperança Garcia

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Prof. Dr. Daniel Jorge Viegas Gonçalves
Prof. Dra. Sandra Pereira Gama

Examination Committee

Chairperson: Prof. Dr. Someone
Supervisor: Prof. Dr. Daniel Jorge Viegas Gonçalves
Member of the Committee: Prof. Dr. Someone

October 2016

Acknowledgments

Agradecimentos.

Todos aqui.

Abstract

English Abstract goes here.

Keywords: one, two, three, keywords

Resumo

Resumo Português fica aqui.

Palavras-chave: uma, duas, três, keywords

Contents

List of Tables

List of Figures

Acronyms

CMYK Cyan, Magenta, Yellow, Key. 13, 14

CSV Comma Separate Values. 11, 12, 17

HSV Hue, Saturation, Value. 13, 14

InfoVis Information Visualization. 17

IST Instituto Superior Técnico. 10

RGB Red, Green, Blue. 13–15

RNL Rede Novas Licenciaturas. 10

RYB Red, Yellow, Blue. 28

Chapter 1

Introduction

Introduction goes here.

... USAR ESTA PARTE DA INTRODUÇÃO DO RESEARCH PROPOSAL ...

In this section, we introduce the majority of topics to be further studied, the different phases of our research, the metrics we are going to collect and how we are going to treat them. Since we aim to *study to what extent can a user distinguish different amounts of blended colors, when using color mixtures to convey information*, it is important learn from previous results, testing out not only the validity of them but also some missed opportunities.

There are several aspects to be considered when developing the broadest study possible: regarding color blending profiling tests, it exists - among others - some questions which remain unanswered; some of them were risen in the studies by Gama and Gonçalves [GG14b, GG14a]. These questions can be divided in four categories:

1.1 Dissertation Outline

Describe the organization of the dissertation document, referring to other chapters.

Chapter 2

Background

2.1 Theoretical Background

2.1.1 Color Perception

2.1.2 Color Models and Spaces

2.2 Related Work

2.2.1 Color Blending Research and Techniques

Ver se houve desenvolvimentos na área em 2016.

2.2.2 User Color Studies Online

Ler artigos do David Flatla, investigar se existe trabalho feito na área nos últimos meses.
Procurar artigos do CHI.

2.3 Discussion

Discutir aqui possíveis questões que podem ser abordadas.

- **Questions raised Before**
  - Will perceived colors correspond to a particular fixed angular value, in the color wheel?
  - Which is the best formula to blend colors, in each color model? Is it linear interpolation or another?
  - In the case in which 3 colors are blended, do observers realize all colors at the same time or do they decompose the mixture, firstly in a mixture of two colors and then a blending of a third color?
  - What is the best way to present color, without influencing color perception?
  - Does the user *really* understands which colors are involved in a mixing?
- **Perception Questions**

  - Does the order in which colors are mixed, influence mental mixing models? Are there common patterns among mixing orders?
  - Do shapes and proximity, influence how color is perceived?
  - Until which extent does background influence the perception of a subject, in particular a blended color?
  - If color parameters like Saturation, Value or Luminance change in a blending, does it modify color blending perception?

- **Information Visualization Questions**
  - Do continuous scales yield better results than discrete color scales?
  - What is the influence of nominal color scales in perception?
  - What are the results if no color scale is presented to guide the user?

- **Cultural Questions**
  - Does the gender really influences how the color is perceived? Is it possible to observe a significant gap between male and female answers?
  - Is it possible to observe significant differences in observation, depending on user's cultural background?

Although there are these questions whose answers remains to be found, only a portion of them will meet their answers, since this is Master Thesis Research Problem. However, there is a set of these questions which was considered crucial and, consequently, had more priority above others: it was this set which was the focus of our studies. We intended to perform **three studies** and, in the following sections, it is covered the entire proposal for the first study, the conditions in which the study is ideally performed and other important details.

Chapter 3

BlendMe!

3.1 Objectives

Justificar nome
Remember the objectives.
Referir HSL com hue variante, e SL com parametros maximos.
Estabelecer questões claras e perguntas que necessitam de resposta, para nos resultados serem mapeados com respostas.
Tipo: Q1: ...

APROVEITAR...

As previously referred, only a set of questions is answered in our research. The **goals for the User Study** are as follows:

- Study one way to present color which does not influence color perception.
- Understand if there is any psychological organization of color, when detailing color mixtures' components.
- If possible, obtain results to ascertain the cultural influence in color perception.
- Study the influence of discrete and continuous color scales.

Additionally, it is relevant to understand which color model stands as the best to mix colors which are, from a perceptual point of view, more similar to the users expectation.

We have planned to develop this study in three different strands: in a **Laboratory Environment** (which will allow us to calibrate and perfectly control the entire study conditions), in an **Online Environment** (which will allow us to disseminate our study to a larger set of users, even without controlling the calibration of the testing environment) and, finally, using **Mechanical Turk Environment** (Amazon's worldwide crowdsourcing marketplace to perform Human Intelligence Tasks, which we would use in order to acquire a huge set of users, even though we can be dealing with speed-clicker users and letting go almost all environmental control).

To meet these study requirements, we drafted our study into three different phases: a **User Profiling Phase**, a **Calibration Phase**, a **Color Deficiency Test Phase** and finally, the **Core Phase**. In the following sections, we detail each of these sections.

3.2 Designing the Solution

Design the implementation, talk about the process ever since wireframing, through the mapping of concepts between what we want and how we implemented, in order to achieve what we want. Include screenshots from the implementation.

Important detail: color conversion between Excel and adapted colors with ICC profile, Spyder and all. ColorConverter.m.
Dividir secção em partes do estudo, introduzindo com Research Proposal para motivar decisões. Referir todos os detalhes de implementação. Justificar completamente todas as decisões que foram tomadas (número de placas, informações pedidas, métricas colhidas, tudo.) Falar de folha de calculo do excel com todas as cores, que depois for migrada para matlab e convertida de acordo com perfil de calibração.

3.2.1 User Profiling Phase

FALTA TEXTO AQUI...

In this phase of the study, questions about the Age, Gender, Academic Degree, Nationality, Country of Residence and the Native Language of the user should be asked: these questions will help us conceiving user profiles with key indicators about cultural background and gender relation to results of each test.

(... ACRESCENTAR AQUI MAIS INFOS RELEVANTES ...)

When this phase is concluded, the user will be guided to another stage of the study, to perform the **Calibration Phase**, where he will be asked to analyze a set of images and answer a pair of questions.

3.2.2 Testing Calibration Phase

Performing online tests - specially when trying to obtain precise values about color - carries obvious problems on how it is guaranteed the results which may appear are, in fact, compliant with certain patterns of quality, specially color and monitor calibration patterns.

To overcome this problem, the ideal solution is to develop a system capable of acquiring information about the user's monitor calibration, *e.g.* Brightness, Contrast, RGB Color Balance, Gamma or Saturation, as a pre-step of the study and apply an appropriate calibration when rendering the study's main page. Since we have not found a way to tackle this solution so far, we developed another solution: to present, as pre-step, some calibration images in which the user will have to perform a set of small tasks, indicating us a set of answers; in the end of the test, we have to analyze the answers to verify if they are compliant with a certain pattern of calibration acceptability, determining if the answers of a certain user can be considered true and not misleading.

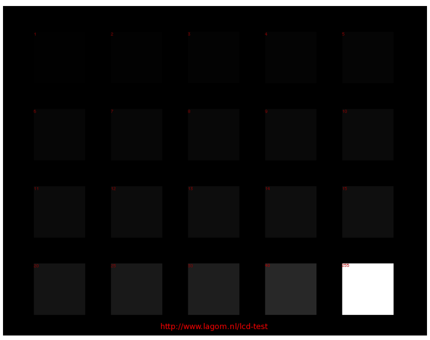

Figure 3.1: Example of Black Squares Test.[1]

We will present to the user a **pair of images**, consisting of a set of 10 squares each; in the first of them, we will test the black level of user's monitor, showing each square with a different shade of black, from brighter ones (Square 1) to darker ones (Square 10). The user is, then, asked to express which square is the last, most perceivable, darker square, and point out the word which follows it. The second image has the same job as the first one, except that this rules out the white-level of the monitor and the ten squares present different tones of white. An example of this sort of test is presented in Figure 3.1. The results of this phase are recorded for further analysis when the study finishes.

Regarding the laboratory environment, we are going to conduct the user tests in a LCD monitor, under a fixed light source; the monitor will be calibrated using a colorimeter which will consider the existing light in the environment and adjust the color of each pixel to a standard. The user will be focused on the task and no other user will be present in the room at the same time, excluding the study regulator; the user will have a fully detailed test protocol to follow. ***(... FALTA TEXTO AQUI ...)*** Referir limitação descoberta entre Chrome Safari

3.2.3 Testing Color Vision Deficiencies Phase

3.2.4 Core Test Phase

Incluir tabela com todas as cores, igual a folha de auxilio. Referir que Ciano, por erro, não esta a ser testado no formato objTwoColors.
Referir aqui que dados estão a ser guardados do utilizador, e como estão a ser guardados, (objTwoColors e twoColorsObj), etc. Referir aqui também que slider contemplava cores standard da folha de calculo para ambiente online, mas para ambiente laboratorio cores eram antes processadas no Matlab. Slider não tinha cores ordenadas para que utilizador não utilizasse algum modelo mental e aprendesse previamente a misturar. Cores foram misturadas sem qualque critério (referir ordem pela qual apareciam). Referir aqui caminho do calibrador -¿ icc -¿ matlab -¿ converter cores -¿ slider

3.3 Evaluation Criteria

Ishihara plates and more, whatever we consider relevant. Falar também de como a calibração era considerada válida ou não. Erros que poderiam ser gerados pelo field number html5, que com scrolls podia dar valores errados.

3.4 Divulgation

MTurk problems, facebook, Reddit, FacebookAds, FNAC prize money.
Como fizemos o controlo dos users (Analytics). Bridge to next chapter.

[1]"Black Level", Available at: `www.lagom.nl/lcd-test/black.php`. Last accessed on September 7th, 2016.

Chapter 4

Analyzing Color Mixing Perception

In this section, we are going to dive into the results obtained from the user study described on chapter number 3. On the first section, we will clearly explain the test protocol which was followed by the users in the laboratory environment to correctly execute the study; this section will be followed, not only by the description of how the gathered data was treated and cleaned (Section 4.2), but also the transformation of this data using *Matlab* processing tools, in order to prepare it for the statistical scrutiny (Section 4.3). Hereafter, conclusions will be drawn from the study at section 4.4, when trying to find answers to the questions/objectives raised before.

The final section of this chapter will be dedicated to summarize the results and infer important conclusions, implications and guidelines which could be relevant for the InfoVis field of research.

4.1 Protocol

The existence of a test protocol, when performing a User Study is mandatory: without it, the test may not follow a strictly previously defined standard. As written before, this user study was conducted two-pronged: in a laboratory environment and *via* online dissemination channels.

4.1.1 Laboratory Environment

The users were given always the same briefing when they arrived at the user study test site: it was explained the motivation behind the master thesis, the goals which were expected for this phase of the study and what was expected for them to execute. The most important information which was told was that "there was no pre-defined correct and wrong answers to each question, this test was designed to test the general color mixing capabilities of the majority of the users". Besides this information, the user study was self-contained, in the sense that every other relevant information and instruction was in the interface, adapted for each test phase, so it was not given any physical artifacts describing instructions. The instructions were available on two languages, depending on the choice of the user: Portuguese and English.

Before each session of the laboratory environment test-run, a Datacolor Spyder 5 Elite Color Calibrator was USB-connected to the computer and, using the software which is shipped along with it, the computer LCD display was fully calibrated (the software offers the option of recalibrating, the option of checking the calibration and also, the option of fully calibrating the display) by testing the pixel emission when emmiting a particular set of colors. The display was everytime fully calibrated, since the software manisfested an erratic bahaviour when using the other functions: the screen colors were presented in a very warmer/colder color profile tha it was before.

The tests were conducted at, most of the repetitions, in Rede Novas Licenciaturas (RNL) at Instituto Superior Técnico (IST), and fewers times in other locations with similar conditions: this is due to constraints in finding users, so the test site needed to have a (limited) mobility feature. However, the conditions remained the same concerning the illumination, the position of the user and the computer used: a Macbook Air 13' (Mid-2013) was prepared undeneath a fixed incadescent light-source (but slightly deviated from it, to minimize light reflections on screen), the user would sit in front of the laptop, in an almost silent environment. Ideal conditions of this test would be such that the user could be sitting alone in a completely silent room, his head would be always at the same distance from the screen, resting in a head-rest and the LCD display's inclination would be perfectly adjusted to the user's eyes.

4.1.2 Online Environment

Performing the study online, as easily predictable, develops some characteristics which cannot be completly controlled. For the sake of calibration, it was asked the user to perform a set of six calibration easy steps before starting the test, so the online user's screen would be, somehow, in a standardized calibration fashion. The calibration steps which were asked are:

1. If possible, adjust your room lights for a comfortable usage of your device.
2. Avoid reflections on your screen, by diverting the screen from direct sources of light. This step is important, since light reflections can affect visualization of images.
3. To adjust the **Black Point** of your screen, define the <u>Contrast</u> and <u>Brightness</u> of your screen to their maximum.
4. After Step 3, gradually reduce **Brightness** value of your screen, in order to correctly distinguish the squares of each image below [calibration squares images].
5. If possible, define the **Color Temperature** of your screen to 6500 Kelvin Degrees.
6. You are now ready to answer the following questions!

The ideal conditions of this test would be such that we could control and maniupulate the color calibration of online user's LCD display, using a software piece which would acquire important informations from the screen configuration, *e.g.* resolution, white-point, black-point, brightness, among others, digest the values and present the questions from the Core Phase in a completely controled and calibrated window. Further investigation could focus in developing this system.

The users were asked to fill in a profiling questionnaire (as seeen of section 3.2.1), as well as to respond to calibration form (Section 3.2.2). A validated simplified 6-plate Ishihara color blindness test [dAK92] is, then performed (Section 3.2.3), before proceeding onto the 32 questions test-phase, in which the user is asked to slide one(two) circular object(s) placed on top of a bar, to indicate a(the) color(s) which he thought were the correct mixture answers. In the end, the user could leave feedback, by sending a message which would be stored in a Relational Database.

The instructions which were presented in each page can be consulted in Appendix A.

4.2 Data Cleaning

Throughout the user study, we collected a **total amount of four-hundred and seventy-nine (477) users** which interacted with our study and fulfilled, at least, until the Color Vision Deficiencies Test Phase. However, only **two-hundred and sixty-one (259) users went on to the core phase of the**

study, representing **54.29%** of the total amount, giving at least one answer on the set of 32 questions, loosing the other two hundred and eighteen (218) users which did not leave any answer, defining the remaining percentage **45.71%.** This large drop of users could be due to errors reported by the users, apparently the inability to submit answers when using the ”Submit” button after rating the question; there were also some complaints when users tried to perform the study in some mobile devices (namely, the *iPhone® 6*), whereon the color slider was not able to be dragged and change the color value at the user will.

Concerning the percentage of users which showed up at the **laboratory trials, there were twenty-nine (28) users who performed the entire study**. On the other hand, there were **two-hundred and thirty-two (231) users which carried out the study online**. There was also a small sample of color vision deficient users which we will analyze in a qualitative manner; this set of users contains only one (1) user from the Laboratory Environment - **3.57%** of the sample size - and two (2) online users - **less than 1%**. Lastly, we detected a small percentage of **six (6) users (2.59%) which did not presented a correct calibration of its LCD display**, evaluation based on the criteria referred before. The data presented and used in this dissertation document was gathered along roughly two months, from 15th of April until 8th of June. As said before, it was collected both with online and laboratory users, which was therefore stored in a Relational Database as previously explained in section 3.2.

In the end of the study, Comma Separate Values (CSV) files were exported from each table using a PostgreSQL for macOS called *Postico*[1], which originated five files containing raw data to be cleaned and processed. The files were as follows:

- *raw_data_user_profile.csv* - Data aggregated from ”Profiling”, ”Calibration” and ”Color Vision Deficiencies” Tables;
- *raw_data_first_profiling.csv* - Data from ”Profiling” Table;
- *raw_data_first_calibration.csv* - Data from ”Calibration” Table;
- *raw_data_first_ishihara.csv* - Data from ”Color Vision Deficiencies” Table;
- *raw_data_first_results.csv* - Data from ”Results” Table;

User ID	Type	First Color	Second Color	Third Color	Drags	Time	Rating	Resets	Question ID
5710cca334d60	objTwoColors	#0080FF	hsl(58.69565217391305,1,0.50)	hsl(98.15217391304348,1,0.50)	992	117	4	2	10
5745c1c07cc0c	objTwoColors	#8000FF	hsl(300,1,0.50)	hsl(324.13043478260875,1,0.50)	645	55	2	1	14
5745350dc1e22	objTwoColors	#0080FF	hsl(226.30434782608697,1,0.50)	NONE	115	11	5	1	10
57451c3b38192	objTwoColors	#00FF80	NONE	hsl(150,1,0.50)	462	39	5	1	15
574511e99b6d9	objTwoColors	#0080FF	hsl(15.652173913043478,1,0.50)	hsl(316.30434782608694,1,0.50)	442	40,	1	1	10
57427cf6bad0c	twoColorsObj	#00FFFF	#FFFF00	#46FF9C	6	14	3	1	32
5740bda9be3dc	objTwoColors	#FF7200	hsl(9.130434782608695,1,0.50)	hsl(50.21739130434783,1,0.50)	45	22	5	1	11
573c783748e8b	twoColorsObj	#00FFFF	#CBFF00	#00FF6B	44	25	3	1	32

Table 4.1: Excerpt of Results Table, with raw data.

The refined tables were then divided into new and more specific ones so that we could detail our results analysis according to the goals defined before; the ”Results” table was refined into Laboratory Results, Online Results and demographic results: concerning the age, we divided it on Users aged below 20 Years Results, Users aged between 20 and 29 Years Results, Users aged between 30 and 39 Years Results, Users aged between 40 and 49 Years Results, Users aged between 50 and 59 Years Results and Users aged above 60 Years Results. Respecting the division of genders, we created the categories Female Users Results, Male Users Results and Other Gender Users Results. An excerpt of raw data contained in ”Results” table can be found in table 4.1; this allows us to support the explanation of the following steps of the cleaning phase.

[1]Postico - a modern PostgreSQL client for the Mac, Available at: `eggerapps.at/postico/`. Last accessed on September 11th, 2016.

Dividing the results among smaller CSV files was the first step of the cleaning phase: the next checklist represents the detailed path which was followed to fulfill the data cleaning.

- **Remove "hsl(..., 1, 0.50)"** - It was needed to remove the extra information stored in columns *First Color, Second Color* and *Third Color*, since this is redundant because it never varies from entry to entry of the table (remember Section 3.1). These values are the *Saturation* (S) and *Value* (V), primitives of the HSV Color Model used.

- **Format Values** - This step was performed just after the previous one. The value which remains to be formatted is simply the *Hue* (H), which is equal to a very precise position on the coded color slider on the interface; the value was composed of 14 decimal numbers, giving us much more precision than what is, in fact, needed considering that the hue is measured in terms of integer numbers. The number was rounded up to its closest integer number, then. Besides that, there was still one value to be adjusted which was the missing response: *NONE* neeeded to be replaced by 0, to simplify the processing of null answers.

- **Sort Entries** - In order to favour the iteration when processing the data, each line of the "Results" Table was sorted according, firstly to the *Question ID*, and after by *User ID*.

- **Normalize Laboratory Data** - As previously said, to perform the Laboratory Study we used a Spyder Color Calibrator to manage the color representation independently of the environmental conditions of light. Since the Color Profile file generated by the calibrator was used to adapt colors to be presented to the user, those same colors had to be trackbacked to the original color, for the sake of normalization of values. This is specially useful when comparing the results from this environment to the "Online" Results, helping in data processing later.

- **Verify Duplicated Entries** - This step was performed only to ensure that the entries would not have any matching copy. As expected, there were not found any copies.

- **Normalize Profiling Info** - Regarding the "Profiling" data, there was some which was written in Portuguese and other in English, depending on the language to perform the study chosen by the user. To avoid misleading profiling categories, all of the academic degrees were normalized to its corresponding name both in English and Portuguese. Also, the raw language values contained some specification of English dialects (*e.g.* en_US, en_UK) and other languages, which was more information than we actually needed; these values were normalized to correspond only to its native and original language (like English, solely).

- **Sanitizing Users** - The tables contained many entries from users that performed the study with incorrect calibration and from users which gave unexpected values on the color deficiencies test phase; the entries which corresponded to a user that failed all 6 values on the later phase, would be deleted, leaving no trace of its participation. Concerning the bad calibration values, it "opened a window" to investigate the resilience of results when the calibration was not what it was expected - this will covered in sub-section 4.5.1. To end up the cleaning phase, it was decided to treat the color deficient users independently: we separated their values from the regular users to perform a qualitative evaluation.

An example of clean data can be found in table 4.2. The next step of data handling is processing it to prepare metrics, establish comparations to pre-calculated answers and depict results in a CIE Chromaticity Diagram. More tables can be found in Appendix C, specifically Section C.1.

User ID	Type	First Color	Second Color	Third Color	Drags	Time	Rating	Resets	Question ID
5713a02a13044	objTwoColors	#00FF00	0	137	459	56	2	0	17
573e4d0eb795b	objTwoColors	#00FF00	235	59	121	28	4	0	17
573edae85268b	objTwoColors	#00FF00	242	57	224	20	5	0	17
5740ad339507d	objTwoColors	#00FF00	228	67	205	14	3	0	17
573c70dabcfe0	objTwoColors	#00FF00	55	221	192	14	2	0	17
57582b17cd76a	twoColorsObj	#FF0000	#00FF00	#AF0049	724	65	2	0	18
573c783748e8b	twoColorsObj	#FF0000	#00FF00	#BFBE00	656	47	3	0	18
573e4022949b1	twoColorsObj	#FF0000	#00FF00	#B000FF	334	23	2	0	18
571151812791a	twoColorsObj	#FF0000	#00FF00	#C9B2A2	110	39	2	0	18

Table 4.2: Excerpt of Results Table, with clean data.

4.3 Data Processing

Processing the data was an important part of the process, since it was important to prepare the raw data collected and compute additional metrics which could be further analyzed to answer the raised questions. To perform this processing, we decided to implement a set of scripts in *Matlab* which could gauge the dataset of each question, demographic group and subset of users (non-calibrated and color vision deficients).

With this data processing, we intend to verify each answer-pair given by a certain user and compare the pairs with each other. It was important to separate the results by question ID, compare each questions' results with other questions that could conceive the same results, blend the values to check which color model answers are closer to (either Hue, Saturation, Value (HSV), Red, Green, Blue (RGB), Cyan, Magenta, Yellow, Key (CMYK), CIE-L*a*b* or CIE-L*C*h*) and also, give meaning to each value, attributing a name to each color. All these parameters and computations are described in the next two sub-sections.

4.3.1 Data Preparation

Given the fact that questions had some differences between each other, there would have to be a cautious analysis; to achieve this, we developed a script for each question, each of file contains the particular set of characteristics ans specific comparisons and values of each question. An exemplary structure of these files can be found on pseudo-code box above. Each question file is capable of computing the following datasets:

- Laboratory Results (Regular Users);
- Laboratory Results (Daltonic Users);
- Online Results (Regular Users);
- Online Results (Daltonic Users);
- Online Results (Uncalibrated Users);
- Demographic Groups: Users Aged Below 20 Years Results;
- Demographic Groups: Users Aged Betweeen 20 and 29 Years Results;
- Demographic Groups: Users Aged Betweeen 30 and 39 Years Results;
- Demographic Groups: Users Aged Betweeen 40 and 49 Years Results;
- Demographic Groups: Users Aged Betweeen 50 and 59 Years Results;
- Demographic Groups: Users Aged Above 60 Years Results;
- Demographic Groups: Female Users Results;
- Demographic Groups: Male Users Results;
- Demographic Groups: Other Gender Users Results;
- Demographic Groups: White Answers (this computation is only available for Questions 1 to 17).

All these datasets are analyzed by a block of code similar to the one in box below __ ; all iterations over each dataset start by verifiying if any value contained in the answer pair is a white (*i.e.* zero valued) answer: if it is, it is stored in a different table, along with all white answers. This was executed **only with non-daltonic users and calibrated users and it was no applied to any type of demographic group**, since its analysis is out of the scope of this thesis. This analysis is interesting, since we can understand if the users opted to leave one value as 0 to truly indicate a white color (to blend and create a lighter color), or simply because they didn't know what to blend.

INCLUIR BLOCO DE PSEUDO CODIGO de cada ciclo Since the colors obtained in the color slider indicate values for the HSV Color Model, it was mandatory to convert the color to a common color standard: for that reason, the values were converted from HSV to CIE-XYZ Color Model. Thus, we can produce color blends in every studied color model (HSV, RGB, CMYK, CIE-L*a*b* and CIE L*C*h*) and ensure that colors obey to the same common standard; also, this is specially important to produce Chromaticity Diagrams where colors are mapped according to a set of XYZ primitives. Both of the answers were blended according to each color model referred before: for models which contained no angular values (RGB, CMYK and CIE-L*a*b*) it was only needed to interpolate the values for each primitive; but for models that have angular values (HSV and CIE-L*C*h have their Hue's value), it was needed to calculate the angular interpolation of their primitives.

$$
\begin{aligned}
R_{final} &= \frac{|R_{C1} - R_{C2}|}{2} + min(R_{C1}, R_{C2}); \\
G_{final} &= \frac{|G_{C1} - G_{C2}|}{2} + min(G_{C1}, G_{C2}); \\
B_{final} &= \frac{|B_{C1} - B_{C2}|}{2} + min(B_{C1}, B_{C2});
\end{aligned} \tag{4.1}
$$

An example of linear interpolation between primitives can be seen on Equation 4.1, in which we blend RGB primitives. The listing 4.1 shows how the angular interpolation is being calculated with our *Matlab* script.

Listing 4.1: Excerpt of *Matlab* code which interpolates the angular Hue value.

```
diff_angles = abs(Hue_C1 - Hue_C2);
if diff_angles > 180
    angle_small = (360 - diff_angles);
    sum_major = max([Hue_C1 Hue_C2]) + (angle_small / 2));
    if sum_major > 360
        hue_final = rem((max([Hue_C1 Hue_C2]) + (angle_small / 2))), 360);
    else
        hue_final = max([Hue_C1 Hue_C2]) + (angle_small / 2));
    end
else
    hue_final = min([Hue_C1 Hue_C2]) + (diff_angles / 2);
end
```

Afterwards, every resulting blending is **compared to the pre-calculated value for each color model**; the distance to the late value is stored for statistical analysis. It is also **calculated the distance to the expected HSV values**, since the colors presented to the user were too calculated in HSV Color Model. Additional comparisons are: the colors blended in HSV are **compared to the expected color pairs of other questions which generate the same (or roughly) color**, to understand if our users tended to mix other pairs than the one expected for that question. To end this comparisons, the centroids of each set of colors mixed in every color model are calculated, along with its distance to the expected pre-calculated answer.

This computation is applied to all questions from 1 to 17, when the resultant color is given and two answers expected. However, there are some differences when processing data from questions 18 to 32, which are questions where the two primitive colors are given and the resulting color is expected, which implies a much simpler analysis due to only have to process one answer: there are no white answers to process (the ones which exist are excluded) and no colors to mix with each other; it is only calculated the distance of the answered color, to the expected one.

4.3.2 Color Bins Comparation

This analysis phase had a very important step, which was to assign meaning to the answers given by the users: **to attribute names to colors indicated**, whose to be commonly used by the users. Ideally, we would conduct a separated user study to perceive which names people normaly attribute to colors; then we would gather all the data and analyze which were the most common names.

Luckily, the web page *XKCD*[2] had already conducted a widely large Color Survey[3] to study wich were the most common RGB color triples among users. They performed roughly more than 222 000 user sessions to ascertain color naming: they produced a map which shows the dominant names attributed to RGB colors over the faces of RGB cube (Figure 4.4), and they also produced a huge file comprised of 196 608 named RGB triplets, grouped by Color Bins.

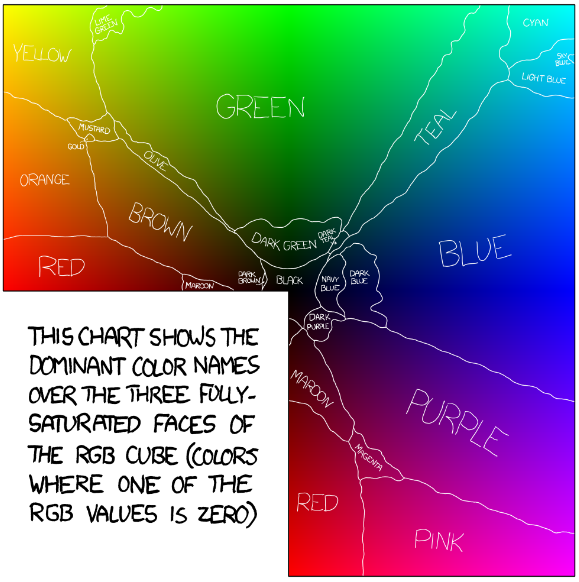

Figure 4.1: XKCD Color Survey: map of color dominant names.[4]

Despite the fact XKCD's Color Survey was not a research realized with a scientific purpose, we decided to use it since it has plenty of information available to compare our results to, and it was executed with a great amount of users which can verify it. We also found great compatibility between this survey and ours, since the values in the first were also presented in its maximum value of saturation (similar to our user study, in which we present colors in maximum hue).

As said, these represent RGB triplets but, being our user results all in accordance with the CIE XYZ Color Model, we needed some cleaning and processing to match the data our responses. We converted the RGB to CIE XYZ values and, then, divided all of the Color Bins in different tables; there are 27 names

[2]XKCD - Stick-figure strip featuring humour about technology, science, mathematics and relationships, by Randall Munroe. Available at: xkcd.com/. Last accessed on September 11th, 2016.
[3]Color Survey Results. Available at: https://blog.xkcd.com/2010/05/03/color-survey-results/. Last accessed on September 11th, 2016.

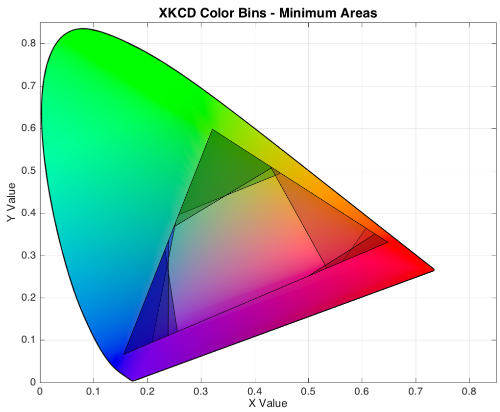

Figure 4.2: XKCD Color Survey: Color Bins Minimum Areas.

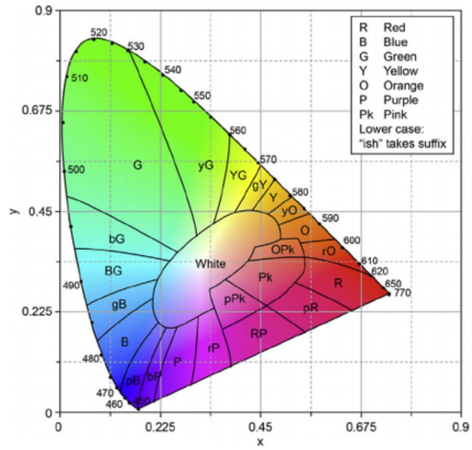

Figure 4.3: Approximate Color Regions on CIE 1931 Chromaticity Diagram. [FM97]

atributted to colors, some have more triplets (*e.g.* blue, green or purple), but some have a smaller set, which could mean greater agreement to assign names to colors. These bins of color are represent with their frequencies, in table 4.3.

Color Bin	Frequency	Color Bin	Frequency
Black	1782	Lime-Green	878
Blue	37725	Magenta	990
Brown	10499	Maroon	3283
Cyan	2625	Mustard	711
Dark-Blue	2233	Navy-Blue	922
Dark-Brown	30	Olive	1336
Dark-Green	2927	Orange	9152
Dark-Purple	669	Pink	12627
Dark-Red	2	Purple	25747
Dark-Teal	163	Red	15474
Gold	49	Sky-Blue	32
Green	47858	Teal	9007
Light-Blue	2078	Yellow	7808
Light-Green	1		

Table 4.3: XKCD Color Survey: color bins.

The idea was to compare our answers with each color bin, to create a mapping between our users' values and commonly-used names; in order to simplify and speed up the computation of the comparations, each color bin was drawn and the lowest polygon formed by the set of triplets of each bin was used to compare the values (instead of comparing each answer with every triplet). Moreover, when processing these sets of RGB triplets, we left 3 color bins out of the game: *Black* since it has no expressivity in the Chromaticity Diagram, *Dark-Red* and *Light-Green* because they have very few triplets to be drawn.

However, we expected that when these color bins were drawn on a chromaticity diagram, they would create independent and more comprehensive shapes: we found that there are overlapping values for some color bins (Figure 4.2), which ultimately complicates the analysis because there is more than one possible name for the same color. For example, *Blue* and *Dark-Blue* share 5 triplets: (0,0,76), (0,0,77), (0,0,78), (0,0,79) and (0,0,80); *Green* and *Dark-Green* share only 1 triplet, (0,57,0). No value was excluded from any color bin, instead we solve the problem by allowing the program to find only one of the names and look no more after finding it. Moreover, the shapes created by each color bin can be

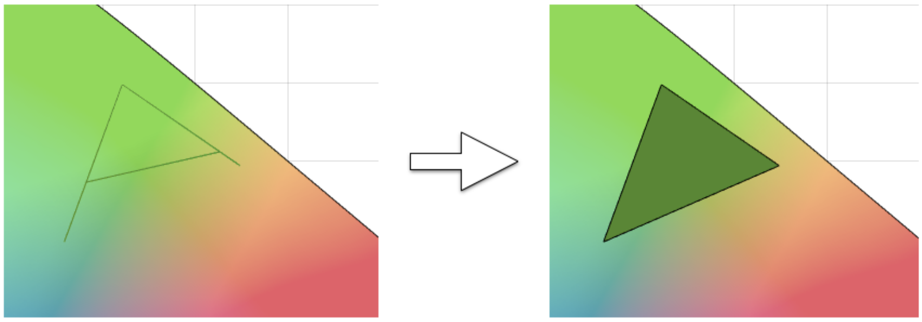

Figure 4.4: XKCD Color Survey: Color Bins Transformation from line of points, to minimum polygon.

depicted as **lines** if they present contiguous values in the same edge of the triangle, or as **triangles** if the values are near the corner of the triangle and are distributed along two edges. The late situation is represented in Figure **??**

Alternatively, we could have opted for another kind of color name detection, but its implementation would be out of the scope of this Master thesis and would take a much longer analysis than the one performed. For instance, one could have gone for identifying the colors by analyzing their values and processing the wavelength originated, comparing the resulting value with calculated areas, roughly defined by Brand Fortner [FM97] and represented on Figure 4.3. This idea had one problem, which was the longevity of its solution (dated from 1997), and the inexistence of a ready-to-go implementation, completed with an svg with all areas defined, had added weight on the decision.

Another possible method to decode color into name would be to interpret color values and associate a color temperature, being the late one compared against a well defined table of values; the problem here was the non-existence of table of well defined values of color temperature. Further investigation is needed to ascertain the possibility of creating such table.

4.3.3 Outputs Generated

Each script ends its execution when saving all the outputs contained in table 4.4: it generates a set of CSV tables ready for being analyzed by SPSS Program, besides creating a great amount of CIE chromaticity diagrams to support the analysis. As referred before, questions 18 to 32 do not generate any output about white answers.

4.4 Results

In the following section, we perform a statistical analysis of the processed data obtained from the user study. The main data to be analyzed is the one obtained in the laboratory environment, being the online data only the corroboration of the main data. We start by drawing a profile of users who responded the survey, characterizing them as the age, country of residence, academic degree, along with other characteristics; after, we begin mapping answers to the questions raised in the beginning of study, also referred in section 3.1, which comprises topics about *Color Mixtures*, *Color Models*, *Color Naming* and consider differences between *demographic groups*.

By the end of this chapter, we are going to discuss the results of the analysis and consider eventual implications for Information Visualization (InfoVis) field of research. The following table **??** summarizes how the study is characterized: how many responses were given *per* question, from which environment

Tables	Diagrams
age_20_results	For each Demographic Group:
age_20_29_results	- RGB, HSV and CMYK Responses
age_30_39_results	- CIE-L*C*h* and L*a*b* Responses
age_40_49_results	For Laboratory Results:
age_50_59_results	- Regular and Daltonic Responses
age_60_results	- RGB, HSV and CMYK Responses
gender_female_results	- CIE-L*C*h* and L*a*b* Responses
gender_male_results	For Online Results:
gender_other_results	- Regular and Daltonic Responses
lab_regular_results	- Uncalibrated Responses
lab_daltonic_results	- RGB, HSV and CMYK Responses
online_regular_results	- CIE-L*C*h* and L*a*b* Responses
online_daltonic_results	For White Answers:
online_uncalibrated_results	- White Responses
white_answers	

Table 4.4: Generated Outputs of Data Processing Phase.

they came and the user sample from each demographic group.

INCLUIR TABELA COM CONTAGEM.

4.4.1 User Profile

As previosuly said on section 4.2, we gathered a total amount of 259 users with, at least, one valid answer: from the laboratory environment we collected 28 users, and 231 from the online strand. All of these users gave valid answers along the Profiling, Calibration and Color Deficiencies Tests Phases: the information collected in the Profiling page is the most important to compose a user profile.

Recalling section 3.2.1, we stablished that the most important informations to collect were the *age*, the *gender*, *academic degree*, *nationality* and *country of residence*, as well as the *native language* of each user. The table 4.5 represents the frequencies of genders, ages and academic degrees.

Environment	Users	Ages						Gender			Academic Degree					
		[0; 20[	[20; 29]	[30;39]	[40;49]	[50;59]	[60;90]	Female	Male	Other	College	High-School	Bachelor	Master	Doctor	NoDegree
Laboratory	28	0	17	5	3	1	2	10	18	0	0	5	13	10	0	0
Online	231	38	145	15	16	11	6	95	134	2	38	42	79	64	5	3
Total	259	38	162	20	19	12	8	105	152	2	38	47	92	74	5	3

Table 4.5: Results: Profiling Information (Gender and Academic)

As seen above, our user sample is composed by 259 (100%) users, being **105 (40.5%) Females**, **152 (58.7%) Males** and a minority of **2 (0.8%) Other gendered users**: this sample age can be characterized as being generally young ($\mu =$ 29.77, $\overline{X_{Age}} =$ 23, $\sigma =$ 40.30), surprinsigly having **8 users (3.09%) aged above 60 years old** which could enhance some interesting differences between age groups. Generally, our users have hight academic qualifications, representing **66.02% of all users (Bachelor, Master and Doctoral Degrees)**, being **38 (14.67%) users qualified with College degree**, **47 (18.15%) have a High-School degree** and only **3 (1.16%) subjects do not presented any academic degree**. Between the laboratory and online environment, the distribution of users remains with the same proportions: more male users than females, mostly aged between 20 and 29 years old (60.71%) and the majority having a superior academic degree (46.43% BSc and 35.71% MSc).

The table 4.6 depicts nationalities, current countries of residence and native languages spoken by our users. From the 259 (100%) participant users, **215 (83.01%) of them have Portuguese nationality**, **217 (83.78%) live in Portugal** currently and **216 (83.40%) speak Portuguese**; the second most

Environment	Users	Nacionality						Country of Residence						Languages		
		CA	ES	PT	UK	US	Others	CA	DE	GB	PT	US	Others	PT	EN	Others
Laboratory	28	0	0	27	1	0	0	0	0	0	28	0	0	28	0	0
Online	231	5	3	188	6	11	18	5	3	9	189	11	14	188	29	14
Total	259	5	3	215	7	11	18	5	3	9	217	11	14	216	29	14

Table 4.6: Results: Profiling Information (Nationalities, Countries of Residence and Languages)

influent group of users are from english-speaking countries (United Kingdom, United States of America, and others). Other minor users which contributed to our survey came from Turkey, France, New Zealand, Sweden or even Antarctica (among others) - **these countries represent only 5.41%**. An ideal distribution of users would be such that it included enough users from all continents, which would give us room to investigate better the cultural implications on the results; another interesting aspect would be to have the users of non-industrialized countries included in the sample, which was not accomplished in this study since an isolated vietnamese user (0.39% of the user sample) contributed to the study.

4.4.2 Color Models

In this section, we will start by analyzing the results from each question, comparing the statistics collected from each color model. Then, we are going to decompose the results by color models and evaluate which questions had the best and worst results. Along with the statistics we are going to present, in the end we are going to map them against the questions exposed before, clearly identifying them whenever they are answered. We would like to emphasize that the important results are the ones collected in the laboratory environment: the online results will bridge and support the conclusions extracted from the laboratory conditions. The values presented as ”distances” are measured inside the finite interval between zero and one ($[0; 1]$), since the Chroamticity Diagram has its values comprised between this interval.

As referred before, each answer pair given from our users on questions 1 to 17 (one color given, two colors asked) was blendend in 5 color models: HSV, CIE-L*C*h*, CMYK, RGB and CIE-L*a*b*. Then, the XY coordinates of each resulting mixture were mapped on a CIE Chromaticity Diagram, the centroids of each group of mixtures was calculated, and the distance from each centroid to the (ideal) pre-calculated answer was measured. The table 4.7 represents the distance measured associated to each color model, for every laboratory result; the table which follows 4.7 is table 4.8, which contains the descriptive statistics for the variables of the first one.

Analyzing Questions

The importance of breaking down the evaluation into questions is such that, it is relevant to understand, at least, which color models have the best results for each mixture. The results from this section will be critical to determine later which color models are more compatible with each mixture, which model yields better results and others who do not. A crucial reminder is that each question set for this user study is blended in HSV Color Model: the primitives anticipated for the blends are, then, colors required to produce the result according to the HSV Color Model.

Question One This is the only question in the entire study which has yellow as the resulting color of the blending. The awaited colors are Red and Green, two primitives from RGB Color Model. Table 4.9 shows the expected colors when mixing in each color model: HSV, CIE-L*C*h*, CMYK, RGB and CIE-L*a*b*.

Question ID	Given Color	Expected Colors		HSV	CIE-L*C*h*	CMYK	RGB	CIE-L*a*b*
1	#FFFF00	Red	Green	0.06	0.17	0.06	0.08	0.09
2	#FF00FF	Red	Blue	0.09	0.13	0.06	0.09	0.09
3	#80FF00	Red	Cyan	0.04	0.22	0.04	0.06	0.07
4	#7F00FF	Red	Cyan	0.07	0.18	0.19	0.24	0.18
5	#FF0080	Red	Magenta	0.09	0.12	0.12	0.1	0.1
6	#FF8000	Red	Yellow	0.05	**0.1**	0.05	0.07	**0.04**
7	#0000FF	Cyan	Magenta	0.15	0.18	0.05	0.03	**0.04**
8	#FF0000	Magenta	Yellow	0.08	0.15	0.03	**0.02**	0.05
9	#00FF80	Green	Cyan	0.05	0.15	0.09	0.09	0.1
10	#0080FF	Green	Magenta	0.23	0.25	**0.01**	0.03	0.05
11	#FF8000	Green	Magenta	**0.03**	0.13	0.13	0.27	0.13
12	#80FF00	Green	Yellow	0.05	0.23	0.12	0.14	0.16
13	#0080FF	Blue	Cyan	0.17	0.14	0.18	0.18	0.12
14	#8000FF	Blue	Magenta	0.1	0.27	0.06	0.13	0.09
15	#00FF80	Blue	Yellow	0.06	0.3	0.5	0.09	0.11
16	#FF007F	Blue	Yellow	0.1	0.2	0.07	0.08	**0.04**
17	#00FF00	Cyan	Yellow	0.05	0.16	0.03	0.08	0.07

Table 4.7: Laboratory Results: Centroids of Results Mixed in each Color Model, for each question, with the distance from itself to the ideal pre-calculated answer. In bold, the best result for each color model.

Statistic	Laboratory (Regular Users)					Online (Regular Users)				
	HSV	LCh	CMYK	RGB	Lab	HSV	LCh	CMYK	RGB	Lab
Type of Distribution	Other	Normal	Other	Other	Normal					
Mean ($\tilde{x}$)	**0.0865**	0.1812	0.1053	0.1047	0.09					
Median (μ)	0.07	0.17	0.06	0.09	0.09					
Maximum	0.23	0.3	0.5	0.27	0.18					
Minimum	0.03	0.1	0.01	0.02	0.04					
Range	0.2	0.2	0.49	0.25	**0.14**					
Stand. Dev. (σ)	0.05267	0.05633	0.11402	0.06947	**0.04153**					
Variance (σ^2)	0.003	0.003	0.013	0.005	**0.002**					

Table 4.8: Laboratory Results: Condensed statistics for centroids of Results Mixed in each Color Model, for each question. In bold, the best result for each descriptive statistic.

Question ID	Given Color	Expected Colors		Possible Results				
		C1	C2	HSV	CIE-L*C*h	CMYK	RGB	CIE-L*a*b*
1	#FFFF00	Red	Green	#FFFF00	#D7A700	#808000	#808000	#C9AB00
Mean Distance to Objective ($\tilde{x}$) - Laboratory				0.13	0.20	**0.09**	0.12	0.12
Mean Distance to Objective ($\tilde{x}$) - Online				0.07	0.21	**0.06**	0.07	0.07

Table 4.9: Question 1, with expected results and mean distances to Objective colors.

As said before, each answer-pair was blended according to those 5 models; after processing data from each question, we performed the mean calculation over each color model, producing the results reported on Table 4.9. One can observe that the model which presents the lowest mean value is the CMYK ($\tilde{x} = 0.09$), possibily indicating that this model yields the best results for this question, while HSV, RGB and CIE-L*a*b* have close values. We performed the Friedman Test in order to evaluate the variance of distances between the blends, and the ideal mixture according to each color model: this test proved ($\chi^2 = 48.568$, $p < 0.05$) that there are statistically significant differences between each color model.

Further analyis with the Wilcoxon Test reveals that CMYK ($p < 0.05$) has, in fact, very different responses from the other color models, which leads us to conclude that **the CMYK Color Model is the best model to represent yellow blends**. Since HSV, RGB and CIE-L*a*b* have similar mean values, we tried to compare them between each other: **there is no statiscally significant differences among HSV, RGB and CIE-L*a*b*, whereby there is insufficient information to evaluate this color models regarding this color mixture**.

It is safe to say that **CIE-L*C*h* has the worst performance against the other color models**, since its mean is the highest of all ($\tilde{x} = 0.20$), and the values from Wilcoxon Test ($p < 0.005$) indicates that this color model has statistically signifcant differences every other model. **This results are corroborated by the online users**.

Question 2 This color question...

Question ID	Given Color	Expected Colors		Possible Results				
		C1	C2	HSV	CIE-L*C*h	CMYK	RGB	CIE-L*a*b*
2	#FF00FF	Red	Blue	#FF00FF	#FB0080	#800080	#800080	#CA0088
Mean Distance to Objective ($\tilde{x}$) - Laboratory				0.49	0.545	0.22	0.119	**0.095**
Mean Distance to Objective ($\tilde{x}$) - Online				0.09	0.17	**0.07**	0.1	0.14

Table 4.10: Question 2, with expected results and mean distances to Objective colors.

Question 3 This color question...

Question ID	Given Color	Expected Colors		Possible Results				
		C1	C2	HSV	CIE-L*C*h	CMYK	RGB	CIE-L*a*b*
3	#80FF00	Red	Cyan	#80FF00	#91C01D	#808080	#808080	#DDA581
Mean Distance to Objective ($\tilde{x}$) - Laboratory				0.49	0.538	0.22	0.119	**0.083**
Mean Distance to Objective ($\tilde{x}$) - Online				**0.04**	0.21	0.06	0.11	0.11

Table 4.11: Question 3, with expected results and mean distances to Objective colors.

Question 4 This color question...

Question ID	Given Color	Expected Colors		Possible Results				
		C1	C2	HSV	CIE-L*C*h	CMYK	RGB	CIE-L*a*b*
3	#7F00FF	Red	Cyan	#7F00FF	-	-	-	-
Mean Distance to Objective ($\tilde{x}$) - Laboratory				0.49	-	-	-	-
Mean Distance to Objective ($\tilde{x}$) - Online				**0.06**	-	-	-	-

Table 4.12: Question 4, with expected results and mean distances to Objective colors.

Question 5

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

Question 12

Question 13

Question 14

Question 15

Question 16

Question 17

Analyzing Color Models

Analyzing the centroids table for each color model, the one which presented the lowest distances was the CMYK Color Model ($\tilde{x}_{CMYK}$ = 0.06), followed by HSV ($\tilde{x}_{HSV}$ = 0.07), RGB and CIE-L*a*b* with the same values ($\tilde{x}_{RGB}$ = 0.09 and μ_{Lab} = 0.09), being CIE-L*C*h the model which has the highest calculated distances (μ_{LCh} = 0.1812). However, the color model which has the highest variance of answers is the CMYK (σ^2_{CMYK} = 0.013), opposed to the CIE-L*a*b* (σ^2_{Lab} = 0.002): this may mean that, aside from having indicated closer mixtures to the desired one, users scattered more their answers when compared to the late color model.

Question ID	HSV			CIE-L*C*h*			CMYK			RGB			CIE-L*a*b*		
	Mean ($\tilde{x}$)	Median (μ)	Std-Dev (σ)	Mean ($\tilde{x}$)	Median (μ)	Std-Dev (σ)	Mean ($\tilde{x}$)	Median (μ)	Std-Dev (σ)	Mean ($\tilde{x}$)	Median (μ)	Std-Dev (σ)	Mean ($\tilde{x}$)	Median (μ)	Std-Dev (σ)
1	0.1258	0.14	**0.08275**	0.2042	0.18	**0.0644**	0.0926	0.07	0.06118	0.1237	0.11	0.07925	0.1232	**0.11**	0.0807
2	0.2173	0.22	0.13014	0.1573	0.16	0.09445	0.106	0.12	0.05642	0.166	0.14	0.10521	0.156	0.17	0.08399
3	0.0941	0.045	0.12308	0.23	0.215	**0.05855**	0.0609	**0.05**	0.03379	0.1114	**0.095**	0.06034	0.1227	**0.11**	**0.03918**
4	0.1162	0.07	0.12975	0.209	0.18	0.11768	0.1971	0.21	0.07302	0.2595	0.28	0.08975	0.2005	0.22	0.09173
5	0.1683	0.15	0.10113	**0.15**	0.15	0.08	0.1322	0.14	0.06787	0.1389	0.12	0.09106	0.135	0.135	0.07579
6	**0.0741**	**0.025**	0.10418	**0.1332**	**0.085**	0.08671	**0.0577**	**0.03**	0.06164	**0.0777**	**0.03**	0.0973	**0.0509**	**0.01**	0.07374
7	0.1577	0.06	0.20679	0.2323	0.27	0.10038	0.0986	**0.06**	0.07523	0.1455	0.1	0.12835	0.1705	0.16	0.07925
8	0.0954	**0.01**	0.16302	0.17	0.17	0.12845	0.1008	0.09	0.4974	0.1292	0.12	0.08967	0.1292	0.14	0.08986
9	0.1258	0.1	0.0989	0.1642	**0.13**	0.07463	0.0911	0.09	0.04642	**0.1005**	**0.08**	0.07524	**0.1142**	**0.1**	0.06694
10	0.3	0.395	0.15792	0.2557	0.275	0.1164	0.1314	0.12	0.04769	0.2107	0.2	0.06403	0.2043	0.175	0.09581
11	**0.0587**	0.03	**0.08828**	**0.1374**	**0.05**	0.11752	0.1304	0.14	**0.02852**	0.1839	0.2	**0.03986**	0.1443	0.15	**0.02727**
12	0.1067	0.04	0.12639	0.2343	0.22	0.06638	0.1314	0.14	0.03966	0.151	0.14	0.07758	0.1743	0.18	0.07284
13	0.2713	0.325	0.15607	0.1663	0.14	0.09701	0.1875	0.215	0.13097	0.245	0.25	0.15795	0.2269	0.235	0.12224
14	0.1186	0.05	0.014423	0.2995	0.315	0.08936	0.0973	0.09	0.4901	0.1327	0.125	0.13406	0.1336	0.13	0.09444
15	0.09	0.095	**0.03559**	0.3031	0.31	0.07872	**0.0625**	0.065	**0.02696**	0.11	0.11	**0.04705**	0.1256	0.12	0.06229
16	0.2087	0.23	0.10378	0.2187	0.25	0.12171	0.1033	0.1	0.05108	0.1593	0.17	0.06829	0.124	**0.11**	0.07872
17	**0.0683**	**0.02**	0.10228	0.1626	0.14	**0.05902**	**0.0452**	**0.05**	**0.0162**	**0.103**	**0.08**	**0.0574**	**0.113**	**0.11**	**0.05112**

Table 4.13: Laboratory Results: Statistics for distances *per* question, of Results Mixed in each Color Model, for each question. In bold, the three best results for each descriptive statistic.

Breaking down the results according to Color Models *per* question, we obtain the results described in the following sub-subsections.

HSV Color Model Blendings The questions which have shorter distances are (in this order) number 11 (**0.03**), 3 (**0.04**), 6, 9, and 17 (all with **0.05**) and, lastly, question 15 (**0.06**); the means related to each question, which can be seen on table 4.13, corroborate the cited distances. Performing the Friedman Test, we know that there are statistically significant differences among all these questions, since all asymptotic significances for all questions are below 0.05. Post hoc analysis with Wilcoxon test, with a significance value of $p < 0.05$ reveal that there are significant differences between HSV and CIE-L*C*h*, but a fairly agreement between HSV and CMYK ($sig_{HSV-CMYK_3}$ = 0.680, $sig_{HSV-CMYK_6}$ = 0.429), RGB ($sig_{HSV-RGB_6}$ = 0.114, $sig_{HSV-RGB_15}$ = 0.116). Relating HSV and CIE L*a*b* has mixed feelings, since the significance values between each of 2 of these questions do not have a statistically significant differences ($sig_{HSV-Lab_9}$ = 0.517 and $sig_{HSV-Lab_15}$ = 0.054) and other 3 have significant differences below 0.05.

On the other hand, the questions with larger distances are (in this order) number 10 (**0.23**), 13 (**0.17**), 7 (**0.15**), 14 and 16 (both **0.1**), 2 and 5 (both **0.09**) which are, again, corroborated by table 4.13. There are some interesting cases related to this color model: some questions which present exactly the same color, but the aim to test if users can present two different types of blends; for example, Questions 6 and 11 show fairly similar values for this color model, since they both show an orange color, but it was intended for user to indicate diffent mixtures (as seen on table 4.7, on Question 6 should be indicated Red and Yellow, and on Question 11 Green and Magenta). This indicates that our users answered quite the same way on both questions, but the distances from the given answers to the expected ones is substantially different ($\tilde{x}_{distance-6}$ = 0.1868 and $\sigma_{distance-11}$ = 0.2467, $\tilde{x}_{distance-11}$ = 0.6448 and $\sigma_{distance-11}$ = 0.1396), which means that **the most intuitive mixture for the users is mixing Red and**

Yellow to obtain Orange, and not the second case.

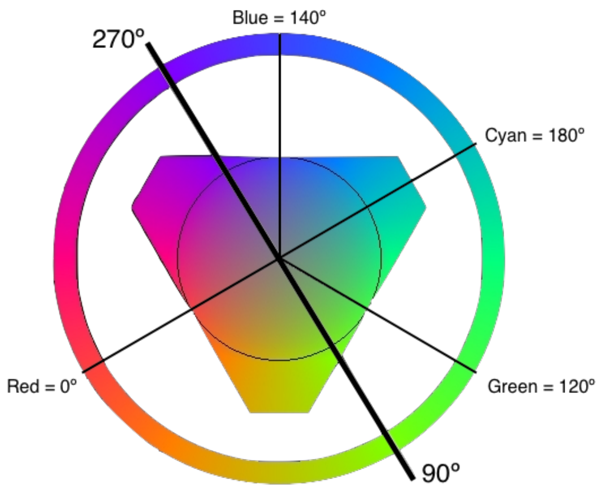

Figure 4.5: Example of HSV Hue Circle, with opposed hue values. Depicted: Question 3 and 4 case, with resulting hues 270° and 90°.

As referred before, there are some color blends that, due to the fact that primitives of the mixture are precisely opposed (180°) in the HSV hue circle, the blending in HSV Model could produce two different outputs. This occurs if one mixes the colors obeying to a positive or negative rotation on the circle (as seen on figure 4.5). There are in this study 3 pairs of colors which blend onto 2 possible colors: pair Red-Cyan (with possible outputs question 3 and 4), pair Green-Magenta (questions 10 and 11) and pair Blue-Yellow (questions 15 and 16). By placing these questions, we intended to understand which side the user would tend to follow (or even if he follows one) when asked to indicate the blending-basis. Therefore, we can produce the following conclusions:

1. Red & Cyan Blend - Analyzing answers for questions 3 and 4, they produce fairly similar results: they are both a bit far from the expected colors ($\sigma_{distance-3} = 0.15736$ and $\sigma_{distance-3} = .023434$), although the mean, median and standard deviation of distances to the HSV precalculated blending, for 3, have a slightly lower value than question 4. This could be also compared to question 12, which also produces **Lime-Green** like question 3, but even when compared, question 3 still has the best results. This leads us to specifiy that, although the results are not great for this mixture, **people have a slight tendency to associate the blending of red and cyan to a result of lime-green, instead of shade of purple**, since the blendings have less dispersion for question 3 (Figure 4.6), than for question 4 (Figure 4.7).

2. Green & Magenta Blend - Similarly to the previous questions, the resuls were quite similar between each other concerning the distance to the expected colors: $\sigma_{distance-10} = 0.15173$ and $\sigma_{distance-11} = .13957$. We can conclude that users **neither *believe* that a shade of blue or orange are produced by Green and Magenta**. As seen before, question 10 produces one of the worst resuls of HSV Color Blendings (along with question 13); instead of believing in a Green-Magenta blending to produce Orange, **the users tend to use again Red and Yellow** as on question 6, which explains the low HSV mean value for question 9, on table 4.13. We can conclude that **the orange color is a strongly implemented color in mental models of our users**. These differences can be seen on Figures 4.8 and 4.8, which largely demonstrates the difference of results dispersion between the two questions, blended in HSV Color Model.

3. Blue & Yellow Blend - Evaluating the results obtained in these two questions (15 and 16), we

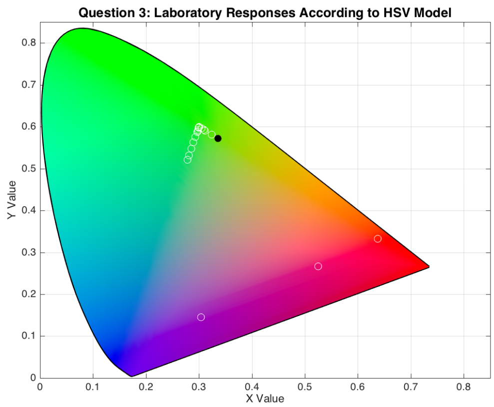

Figure 4.6: Laboratory Results: Answers for Question 3, from regular users, mixed in HSV Color Model.

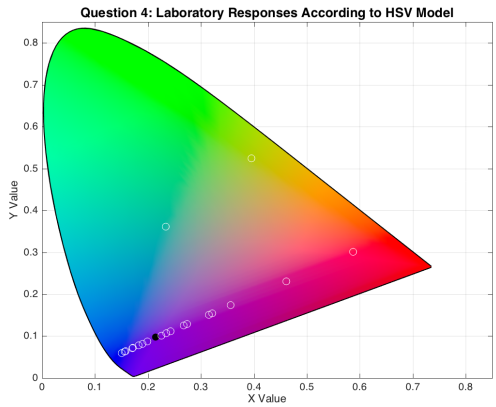

Figure 4.7: Laboratory Results: Answers for Question 4, from regular users, mixed in HSV Color Model.

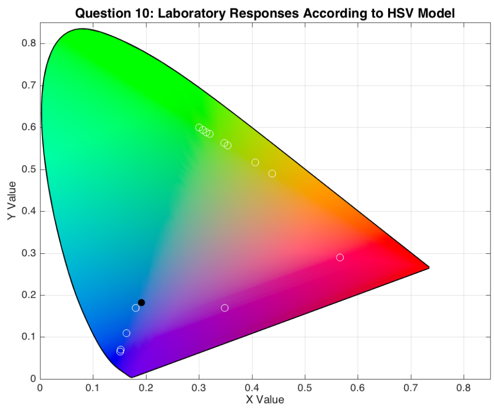

Figure 4.8: Laboratory Results: Answers for Question 10, from regular users, mixed in HSV Color Model.

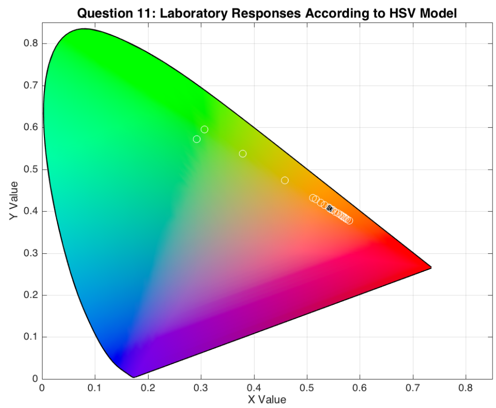

Figure 4.9: Laboratory Results: Answers for Question 11, from regular users, mixed in HSV Color Model.

obtain substantially different results for both. Analyzing results of question 15, we know that the distance to the expect pair of colors does not present the lowest value ($\tilde{x}_{distance-15} =$ 0.6869, $\sigma_{distance-15} =$ 0.2787); however, when blending the result-pairs given on this question, it presents one of the closest distances ($\tilde{x}_{HSV-15} =$ 0.09), while the users did not wander too much when answering ($\sigma_{HSV-15} =$ 0.03559). As it can be seen on Figure 4.10, the resulting blend in HSV Color Model tend to be close to the ideal HSV answer: despite being nearer to the top corner of the HSV Model Triangle, **we can still consider the answers to be a tone of green color**. Regarding the results from question 16, they present themselves to be one of the worst for this color model: the distance to the expected colors is $\sigma_{distance-16} =$ 0.22001 and the standard deviation for the distances of answers blended in HSV is $\sigma_{HSV-16} =$ 0.10378. Since the presented color had a strong value of red component, users tended their answers to the red shade, while blending it with a purple or pink one, to indicate a mixture close to magenta. This indicates that **blending Blue and Yellow, in HSV Color Model, does not correspond to a red shade, but to a green one, according to the users' expectations**.

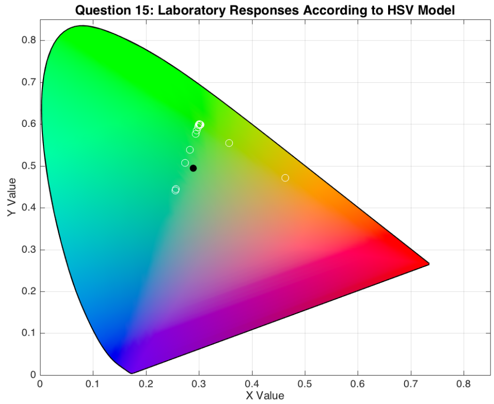

Figure 4.10: Laboratory Results: Answers for Question 15, from regular users, mixed in HSV Color Model.

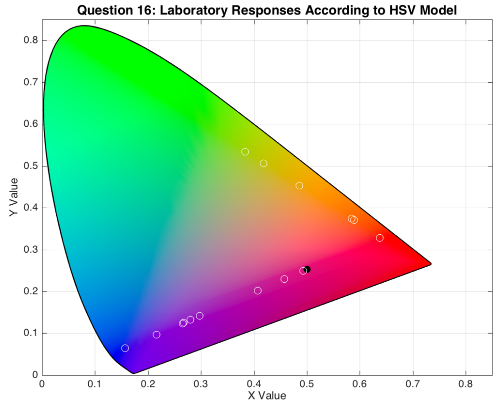

Figure 4.11: Laboratory Results: Answers for Question 16, from regular users, mixed in HSV Color Model.

Another interesting example is Question 10 and 13: they both present a shade of blue, which could be obtained by two strands, either Green and Magenta (Question 10) or Blue and Cyan (Question 13). Particularly, the late question had a fairly simple expected pair (two shades of blue) but the users demonstrated some dispersion when aswering the question: moreover, the statistics prove it, by having a rather high standard deviation ($\sigma_{HSV} =$ 0.15607) when compared to the standard deviation of the whole model ($\sigma_{HSV} =$ 0.05267). These results are, thus, consistent with the know fact [**COLOCAR REF**] that **human color perception is poorer in the blue region of the color spectrum, than in others**, *e.g* the green zone, which produced the best results for this color model; these results are extensible to the purple region of the spectrum, which is contiguous to the blue one. Table 4.14 contains the cited colors and questions, ranked from best (shorter distances) to worst (longer distances) and aims to help the analysis.

	Best					Worst
Shorter Distances	11	3	6	9	17	15
Longer Distances	10	13	7	14	16	2

Table 4.14: Laboratory Results: HSV Centroids, with questions ranked from best result to the worst.

CIE-L*C*h Color Model Blendings From all the color models studied along with this survey, the CIE-L*C*h* model was the one which presented the lower results. The questions which have best results are number 6 (**0.1**), 5 (**0.12**), 2 and 11 (**0.13**), 13 (**0.14**), 8 and 9 (**0.15**); the best reults, by themselves, are quite high when compared with the previous color model, presenting mean distances always above 0.1 (which, in this scale, reveals itself to be a high value). Contrary, question 15, question 14, question 10, question 12 and question 3 have the lowest scores.

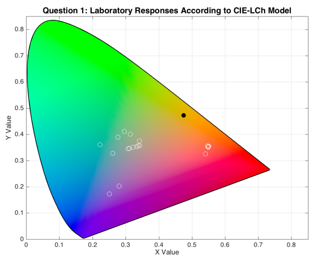

Figure 4.12: Laboratory Results: Answers for Question 1, from regular users, mixed in CIE-L*C*h* Color Model.

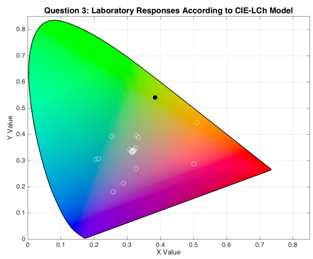

Figure 4.13: Laboratory Results: Answers for Question 3, from regular users, mixed in CIE-L*C*h* Color Model.

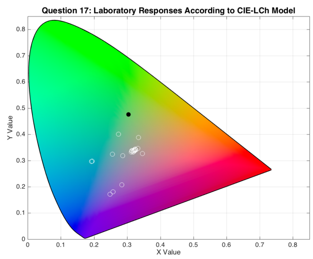

Figure 4.14: Laboratory Results: Answers for Question 17, from regular users, mixed in CIE-L*C*h* Color Model.

The questions which revealed the lowest standard deviations (meaning less osccilation of results) were questions 1 ($\sigma_{LCh-1} = 0.0644$), 3 ($\sigma_{LCh-3} = 0.05855$) and 17 ($\sigma_{LCh-17} = 0.05902$); nonethless, **these questions' centroids are all too far from the ideal L*C*h* expected result**, as seen on Figures 4.12, 4.13 and 4.14.

The ones which **consistently mantained the statistics were question 6 and 11**, having the lowest scores for the mean ($\tilde{x}_{LCh-6} = 0.1332$, $\tilde{x}_{LCh-11} = 0.1374$) which correspond to blends very close to the objective one. However, this values do not beat the ones obtained by the HSV Color Model blends, since their distance is quite longer than the one previously obtained.

Curiously, the best results from this color model are from questions which present red, pink and derived colors like orange, while the weakest results came from a mix of questions which present green shades and blue derived ones, similar to HSV. **For most of the colors presented, this color model should be handled to the user when mixing colors**, since there are other color models with better results than this. The following table outlines the best and the worst answers, ranked in a similar fashion to the previous color model.

	Best					Worst
Shorter Distances	6	5	2	13	8	9
Longer Distances	15	14	10	12	3	4

Table 4.15: Laboratory Results: CIE L*C*h* Centroids, with questions ranked from best result to the worst.

CMYK Color Model Blendings As it was explained before in the theoretical background, this color model is mainly a subtractive one, meaning it will naturally darken the colors when they are blended. The questions with shorter distances are 10 (**0.01**), 8 and 17 (**0.03**), 3 (**0.04**), 6 and 7 (**0.05**) and 1 and 2 (**0.06**), which represent a wide range of colors mixtures. The questions which yielded **better results are majorly related to primitives of the CMYK color model**: Cyan is envolved in 3 blends (questions 3, 7 and 17), Magenta is envolved in 3 blends also (questions 7, 8 and 10), so as Yellow (6, 8 and

17). On the other hand, RGB Model is weakly represented besides Red appearing in 4 questions, with Green appearing in 2 and Blue only in one. This success could be related to the fact that **people tend to formulate mental models of color based on ink mixing in childhood** [GC04], mostly associating it to Red, Yellow, Blue (RYB) and CMYK Color Models without even knowing it.

In fact, when performing the Friedman Test for Variance analysis, we conclude that CMYK Color Model consistently presents the lowest rank against all color models in use, which could **mean greater agreement between the users when indicating their answers**. When comparing it against other color model's results with the Wilcoxon Test, the CMYK color models shows a degree of concordance with the following color models:

- **HSV** - Question 3 ($sig_3 = 0.680$), Question 6 ($sig_6 = 0.429$), Question 7 ($sig_7 = 0.515$), Question 8 ($sig_8 = 0.326$), Question 9 ($sig_9 = 0.551$), Question 12 ($sig_{12} = 0.163$), Question 14 ($sig_{14} = 0.676$) and Question 17 ($sig_{17} = 0.275$).

- **CIE-L*a*b*** - Question 4 ($sig_4 = 0.743$), Question 5 ($sig_5 = 0.739$), Question 6 ($sig_6 = 0.116$), Question 8 ($sig_8 = 0.058$), Question 9 ($sig_9 = 0.112$), Question 12 ($sig_{12} = 0.161$) and Question 16 ($sig_{16} = 0.253$).

- **CIE-L*C*h*** - Question 4 ($sig_4 = 0.444$), Question 5 ($sig_5 = 0.307$), Question 8 ($sig_8 = 0.086$), Question 11 ($sig_{11} = 0.583$), Question 13 ($sig_{13} = 0.959$) and Question 16 ($sig_{16} = 0.130$).

- **RGB** - Question 5 ($sig_5 = 0.536$), Question 8 ($sig_8 = 0.071$), Question 9 ($sig_9 = 0.812$) and Question 14 ($sig_{14} = 0.143$).

The HSV Color Model presents the higher number of questions without a non-statistically significant difference, which means more concordance between the answers of the users; on the other hand, RGB is the least compatible with the CMYK, which is no surprise since it is its complementary color model.

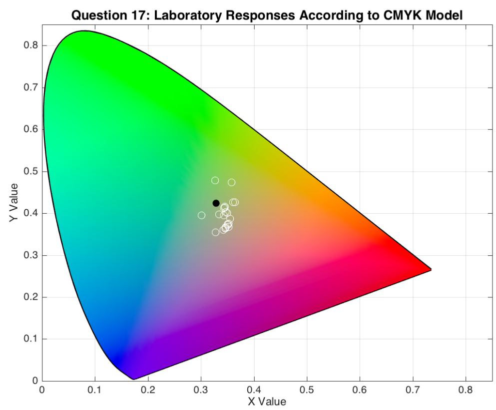

Figure 4.15: Laboratory Results: Answers for Question 17, from regular users, mixed in CMYK Color Model.

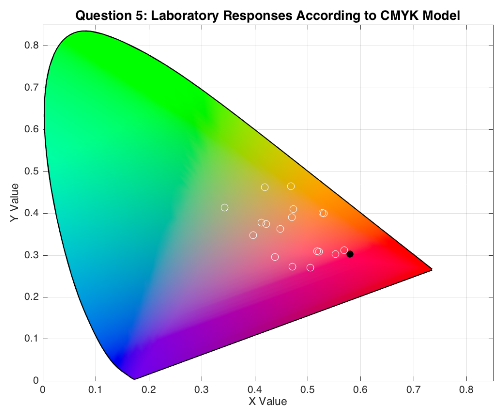

Figure 4.16: Laboratory Results: Answers for Question 5, from regular users, mixed in CMYK Color Model.

The questions with longer distances are 15 (**0.5**), 4 (**0.24**), 13 (**0.18**), 11 (**0.13**) and 12 and 5 (**0.12**), which also represent a wide range of color mixtures. This results indicate that **CMYK is a highly compatible color model with the users expectations**, as many questions contain user answers that are close to the expected pre-calculated CMYK responses. However, this color model is the one which had the most deviation in the distance ($\sigma_{CMYK} = 0.11402$), therefore the most deviation among all answers: we conclude that **for increasing the accuracy of the results and maximize this model's**

results, the number of available colors should be limited mostly to primitive colors like Cyan, Magenta or Yellow, near the most common mental model of color of our users.

Figure 4.15 represents the results of Question 17, which had closer distances to the pre-calculated CMYK blend, and Figure 4.16 show the results for Question 5; this last question has more sparse results, resulting in a more distant centroid. The next table sums up all the questions addressed. The table below gathers the best and worst questions about this color model.

	Best							Worst
Shorter Distances	10	8	17	3	6	7	1	2
Longer Distances	15	4	13	11	12	5		

Table 4.16: Laboratory Results: CMYK Centroids, with questions ranked from best result to the worst.

RGB Color Model Blendings This color model, as explained on the theorical background of this document, is complementary to the CMYK color model. Feedback collected from the users was such that, sometimes, the users which knew how to blend in subtractive color models, tended to be confused and tried to mix additive color models, also. Based on the results from RGB blends' centroid, we can tell that the results from this color model are quite similar to CMYK results: the shorter distances are on questions 8, 7, 10, 6, 3, 1, 16 and 17 while the longer are on questions 11, 4, 13, 12 and 14.

When comparing the results from the RGB color model, the majority of users did reveal lack of knowledge in mixing the colors according to an additive color model, as they tended to mix colors according to the CMYK color model. However, this color model has a high degree of compatibility with every other model, excepting CMYK: RGB presents non-statistically significant differences ($sig_{RGB-x} >$ 0.05) with HSV and CIE-L*a*b* in 10 questions both, and with CIE-L*C*h* in 9 questions. Regarding the CIE-L*a*b* model, the results are particularly dramatic: questions with no relevant differences, the values of the asymptotic significance are quite high, but when analyzing the results from Wilcoxon test for the statistically significant differences, these values are very low ($sig_4 =$ 0.0, $sig_6 =$ 0.0, $sig_9 =$ 0.003, $sig_{11} =$ 0.0, $sig_{12} =$ 0.001, $sig_{15} =$ 0.010 and $sig_{16} =$ 0.004).

Similar to previous color models, question 6 yields the best results when analyzing the mean value ($\tilde{x}_{RGB-6} =$ 0.078), being opposed to question 4 with the highest value ($\tilde{x}_{RGB-4} =$ 0.260). **The orange color continues to be the color with the best results across every model**.

The Figure **??** shows the responses given by the laboratory users to question 8, while the Figure **??** presents the same answer-pairs but blended in RGB Color Model. As it can be seen, results (represened in black empty circles) are a bit scattered, but around the ideal response (black filled circle).

	Best							Worst
Shorter Distances	8	7	10	6	3	1	16	17
Longer Distances	11	4	3	12	14			

Table 4.17: Laboratory Results: RGB Centroids, with questions ranked from best result to the worst.

CIE-L*a*b* Color Model Blendings This color model is the one which has the lowest range of values for the distance of centroids ($range_{Lab} =$ 0.14, wobbling between 0.04 and 0.18). The questions which had shorter distances to the ideal RGB pre-calculated answer are 6, 7, 16 (all with **0.04**), 8 and 10 (both **0.05**), 3 (**0.07**), 1, 2 and 14 (all with **0.09**) and, finally, 5 and 9 (**0.1**). The set of questions with larger distances are: questions 4, 12, 11, 13 and 15.

By analyzing the results from Friedman Variance Test, we conclude that all questions show similar distributions among themselves (mean rank varying between $rank_{min} = 3$ and $rank_{max} = 3.67$), except for questions 6 ($rank_6 = 1.91$), 16 ($rank_{16} = 2.23$), 4 ($rank_4 = 2.57$), 5 ($rank_5 = 2.81$) and 9 ($rank_9 = 2.97$). On the other hand, CIE-L*a*b* has a fairly high degree of compatibility with all other color models: this could be due to the fact that **CIE-L*a*b* is capable of representing all human perceivable colors**, turning this color model closer to the users expectation. When processing the distances from the centroids to the ideal answers, we conclude that **even the "worst" question (number 4, $distance_{Lab} = 0.18$) is, by far, better than the worst questions from other models**, as seen on Table 4.7.

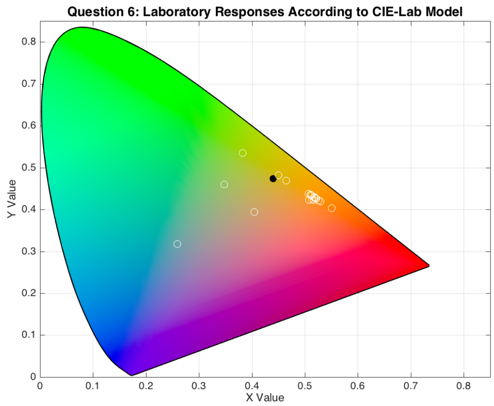

Figure 4.17: Laboratory Results: Answers for Question 6, from regular users, mixed in CIE-L*a*b* Color Model.

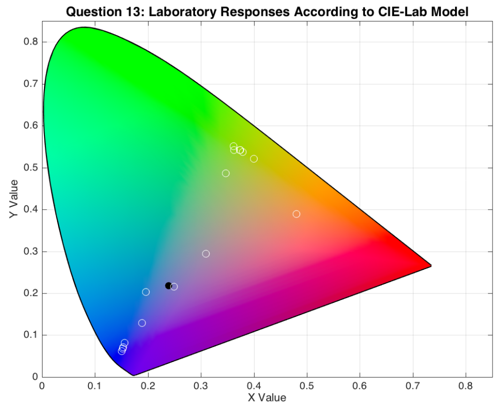

Figure 4.18: Laboratory Results: Answers for Question 5, from regular users, mixed in CIE-L*a*b* Color Model.

The question which has the lowest mean is, again, question number 6 ($\tilde{x}_{Lab-6} = 0.051$) having in this color model its best result, when compared against other color models. Contrarily, question 13 has the most distant mean value ($\tilde{x}_{Lab-13} = 0.227$), once again a shade of blue, consistent with the results from other color models. These questions can be analyzed on Figures 4.17 and 4.18.

	Best									Worst
Shorter Distances	6	7	16	8	10	3	1	2	14	5	9
Longer Distances	4	12	11	13	15						

Table 4.18: Laboratory Results: CIE L*a*b* Centroids, with questions ranked from best result to the worst.

corroborar com online

comparar com tipo oposto de questão.

4.4.3 Color Mixtures

Regarding the color mixtures' related questions, we would like to make the following analysis. We asked at Section 3.1 ...

Ver também se existe cor preferida para começar mistura.

Ver misturas que tiveram maior distancia às cores esperadas. Analisar e perceber se respostas dadas também produzem a mesma cor pedida. Ver misturas que tiveram menor distancia às cores esperadas.

Primary Colors

fazer analise concreta das questões 1, 2, 7, 8 e 17, que são as primitivas. perceber se resultados são positivos para a geração de primitivas com base noutras.

Comparar misturas que originam a mesma cor, com base em primárias diferentes e perceber se utilizadores conseguem detectar várias misturas para uma mesma cor.

Difficulty While Blending Colors

Fazer também mistura mais fácil, comparando os ratings das questões e ver qual a mistura que apresenta melhores resultados.
Falar de respostas em branco, analisar somente valores e perceber se é desconhecimento.

4.4.4 Color Naming

Cores mais comuns em algumas perguntas; existe alguma ordem caracteristica quando utilizador especifica uma mistura?
Comparar se, mais do que pelos valores anteriormente calculados, se users identificam ”uma mistura de vermelho com azul” como dando um resultado magenta, ou mais afinado para uma dada cor.

4.4.5 Demographic Results

Fazer apenas comparação de faixas etárias entre si, e géneros entre si.
Uma ánalise interesse seria comparar faxias etárias por género, mas seria necessária uma amostra bastante mais significativa para cada grupo.

Age Groups

Gender Groups

4.5 Discussion

Fazer apanhado dos resultados todos.
abordar questão de que modelos originaram melhores respostas, se são aditivos ou subtrativos.

4.5.1 Calibration Resiliency

Como verificamos ainda alguns users com calibração imprópria para teste, considerámos que poderia ser uma fonte de resultados interessantes. Como tal, criámos um dataset para os mesmos e comparámos com os resultados dos utilizadores calibrados. Os resultados são os que se seguem...

4.5.2 Creation of Color Scales

Aproveitar resultado de respostas em branco

4.5.3 Color Organization

4.5.4 Implications for InfoVis

Resumo dos resultados todos e regras que se podem levar deste trabalho para a área de InfoVis em geral.

Chapter 5

Conclusion

5.1 Limitiations

Most of the data cleaning would be avoided if the collected were saved in a different fashion (NONES, HSL)...

5.2 Future Work

Software de calibração remota.

Bibliography

[dAK92] Dilogen de Alwis and Chee Kon. A new way to use the ishihara test. *Journal of Neurology*, 239(8):451–454, 1992.

[FM97] Brand Fortner and Theodore E. Meyer. *Number by Colors*. Springer-Verlag New York, 1 edition, 1997.

[GC04] N. Gossett and Baoquan Chen. Paint inspired color mixing and compositing for visualization. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 113–118, 2004.

[GG14a] Sandra Gama and Daniel Gonçalves. Studying the perception of color components' relative amounts in blended colors. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, NordiCHI '14, pages 1015–1018, New York, NY, USA, 2014. ACM.

[GG14b] Sandra Gama and Daniel Gonçalves. Studying Color Blending Perception for Data Visualization. In N. Elmqvist, M. Hlawitschka, and J. Kennedy, editors, *EuroVis - Short Papers*. The Eurographics Association, 2014.

Appendix A

User Study Protocol

A.1 Motivation

By conducting this first study, we intend to:

- Conclude if there is any chance that cultural behaviours influence the user's color perception.
- Realize which color mixtures are more easily perceived by humans.
- Understand if, by using color, it is possible to clearly and easily convey information. This can be particularly interesting and useful when visualizing graphs or maps.
- Conclude if a person is capable of, not only building a mental color mixture model, but also deconstructing mixtures into their basic components.

A.2 User Profiling Phase

This study is anonymous and should take you up to 15 minutes. Please, answer the following answers accordingly.

A.3 Testing Calibration Phase

In this step, it's going to be presented to you a set of images. You should tune you screen definitions, in order to answer the questions, keeping them until the end of this study.

Please, follow the steps below indicated and answer the questions.

1. If possible, adjust your room lights for a comfortable usage of your device.
2. Avoid reflections on your screen, by diverting the screen from direct sources of light. This step is important, since light reflections can affect visualization of images.
3. To adjust the **Black Point** of your screen, define the Contrast and Brightness of your screen to their maximum.
4. After Step 3, gradually reduce **Brightness** value of your screen, in order to correctly distinguish the squares of each image below [calibration squares images].
5. If possible, define the **Color Temperature** of your screen to 6500 Kelvin Degrees.

6. You are now ready to answer the following questions!

NOTE: These 6 steps are only available to the Online Users, since the Laboratory Users do not need to perform these steps as the LCD display is already calibrated.

A.4 Testing Color Vision Deficiences Phase

This is the Color Vision Deficiencies Test.
In this step, it is going to be presented six plates with a colored pattern. Your job is to identify the number present in each plate, typing it down in the text box below. According to your answer, this test will inform us if you have any type of color vision deficiency which may undermine the job of color detection.

A.5 Core Test Phase

Choose the Resulting color which you believe it is the result of mixing the First and Second color, by adjusting the slider below the Resulting color.
Choose two colors with which you can achieve the Resulting color, by adjusting the sliders below each color.
NOTE: These instructions appear alternately, depending on the type of question which is shown.

Appendix B

Chromaticity Diagrams

Include Diagrams generated of CIE Horseshoe.

Appendix C

Tables

C.1 Results Tables

Users
Um exemplo de uma tabela de resultados raw
Um exemplo de uma tabela de resultados clean
Um exemplo de uma tabela de calibração