https://doi.org/10.2352/issn.2694-118X.2020.LIM-22

Evaluation of the Human Visual System in Cosmetics Foundation Colour Selection

Altynay Kadyrova[1], Majid Ansari-Asl[1], Eva Maria Valero Benito[2];
[1]Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway
[2]Department of Optics, University of Granada; Granada, Spain

Abstract

Colour is one of the most important appearance attributes in a variety of fields including both science and industry. The focus of this work is on cosmetics field and specifically on the performance of the human visual system on the selection of foundation makeup colour that best matches with the human skin colour. In many cases, colour evaluations tend to be subjective and vary from person to person thereby producing challenging problems to quantify colour for objective evaluations and measurements. Although many researches have been done on colour quantification in last few decades, to the best of our knowledge, this is the first study to evaluate objectively a consumer's visual system in skin colour matching through a psychophysical experiment under different illuminations exploiting spectral measurements.

In this paper, the experiment setup is discussed and the results from the experiment are presented. The correlation between observers' skin colour evaluations by using PANTONE Skin Tone Guide samples and spectroradiometer is assessed. Moreover, inter and intra observer variability are considered and commented.

The results reveal differences between nine ethnic groups, between two genders, and between the measurements under two illuminants (i.e. D65 and F (fluorescent)). The results further show that skin colour assessment was done better under D65 than under F illuminant. The human visual system was three times worse than instrument in colour matching in terms of colour difference between skin and PANTONE Skin Tone Guide samples. The observers tend to choose lighter, less reddish, and consequently paler colours as the best match to their skin colour. These results have practical applications. They can be used to design, for example, an application for foundation colour selection based on correlation between colour measurements and human visual system based subjective evaluations.

Introduction

The cosmetics field mainly relies on colour attributes as colour is an inherent part of our lives. Colour can distinguish different beauty companies' value in the market in terms of diversity and quality of colour they can offer. More specifically, colour plays a vital role in skin colour matching with the makeup product colour. This study targets specifically the selection of foundation colour.

A foundation can be in cream, liquid and powder forms, and be used as a base for facial makeup. It is the second most popular product [1]. Despite the popularity of this product, choosing the right foundation colour is still a difficult task. For example, Estée Lauder is one of the largest beauty brands and they found in an independent study that 94% of women wear the foundation shade that does not match their skin [1]. Jain et al. mentioned that their communication with beauty companies showed that selection of a shade is a current challenging problem. They conducted an experiment where first participants selected their foundation shade by themselves visually and afterwards experts selected foundation shades for the participants. The participants needed to decide if foundation shade selected by the experts was the same what they targeted. Around 80% of participants preferred foundation shade selected by the experts. As a result, this study revealed that automated consumer beauty application for foundation shade colour selection is very demanded [1].

Various illumination conditions affect the skin colour perception. We focus on skin colour evaluation under two different illuminants (D65 and F (fluorescent)) to test the human visual system's ability to select the right PANTONE Skin Tone Guide sample colour that best matches observers' skin colour through a psychophysical experiment. It is worth mentioning that different ethnicities and genders have wide range of skin colours. In order to measure skin radiance or reflectance, spectroradiometers or spectrophotometers can be used respectively. In our study, we used Photo Research's SpectroScan 745 spectroradiometer. We chose spectroradiometer instead of spectrophotometer because the latter contacts the skin which can change blood flow. Thus, skin colour might vary.

We propose to find the closest PANTONE Skin Tone Guide samples to body parts of individuals using spectral data. Then we compare them to the individual's preferred PANTONE Skin Tone Guide samples. We report the accuracy of the visual system in choosing the right colour of the foundation (i.e. PANTONE Skin Tone Guide sample) in comparison to spectral data as ground truth.

We first describe the related works. Afterwards the methodology, the details of the experiment, and the results obtained with discussion are presented. Finally, the main conclusions are described.

Related works

The appearance of skin has been studied mostly for rendering purposes in computer graphics [2], for face detection and tracking in computer vision [3-4], for diagnostic purposes in dermatology [5], and for makeup and skin care in cosmetics [6]. A challenging problem in skin appearance studies for the applications in cosmetics is to find the right colour for foundation that best matches with the human skin colour.

The human skin colour markedly influenced by the age and health of the people [7]. Although colour is easily perceivable, it is difficult to assess it in an objective way [8]. Moreover, colour is maybe better described as a sensory perception rather than a physical quantity and thus colour data should be very carefully interpreted [8]. The perceived colour information of the human skin can be represented using CIELAB values because there is a potential of CIELAB colour space parameters in cosmetics application [9]. The skin colour is dependent on melanin, variety of blood vessels, degree of the blood oxygenation, and many more parameters [10]. The assortment of different colour complexions is the main concern of the cosmetic chemists to create skin makeup product [10], particularly foundations. There is a common problem in cosmetics that consumers tend to choose foundation colour that does not match their skin colour intentionally. In other words, there is a tendency that some individuals want to look lighter or darker due to various reasons (aesthetics, medical, cultural, etc.) [6].

There is little research done on evaluation of the human visual system in choosing the right foundation colour using objective and scientific approaches. Jain et al. used cosmetics experts as the ground truth to evaluate consumers foundations selection [1] while the experts did the selection subjectively by themselves. Also, skin colour objective measurement has been made mostly by narrowband reflectance spectrophotometers developed specifically for dermatology [11]. The main drawback of spectrophotometers is the area measured is about 0.05 cm^2 but skin is not homogeneous. Additionally, the pressure of the probe on the skin can be an important source of bias. Furthermore, Wang et al. showed that different kinds of measurement instruments measure the skin differently [12].

Hence, our research work addresses two current important problems in cosmetics related to objective evaluation of the accuracy level of the human visual system in skin colour matching for foundation colour selection and efficiency of skin measurement instrument due to non-homogeneity of skin surface. Our study results can be integrated in practical applications for foundation colour selection (e.g. mobile application, guideline for both consultants and customers in beauty stores, etc.).

Methodology

Our methodology consists of a psychophysical experiment with observers and measurements with Photo Research's SpectroScan 745 spectroradiometer for data collection, and usage of RMSE (root mean squared error) and GFC (goodness fit coefficient) [13] metrics for data analysis, and MCDM (mean colour difference with respect to mean) [14] for inter and intra observer variability purposes.

We asked our observers to choose the PANTONE Skin Tone Guide sample colour that best matches with his or her skin colour in order to avoid observers' tendency to choose their preferred skin colour instead.

Experiment

There are different psychophysical methods to use when conducting experiments with observers such as pair comparison, rank order, category judgements, and others. However, we did not use any of those methods. In our case, we gave freedom to the observers to choose the samples among a given samples. The main reason for this is to mimic makeup store conditions.

The experiment was conducted in a dark room in which the only light present came from the light booth cabinet. 110 PANTONE Skin Tone Guide samples were used for the experiment and all 110 samples' radiance values were computed by using the Photo Research's SpectroScan 745 spectroradiometer under D65. Also, the samples were cut into smaller sizes in order to fit light booth cabinet dimensions so that observers can make their choice without any space limitations. Furthermore, for each sample small square holes were made to make easy to compare samples with observer's skin by putting the sample above the specific hand/arm part and compare through the hole. Ten observers from nine countries participated, all of them either university students or academic staff. The observers kept their positions constant.

All observers had normal colour vision and 50% of them used their glasses during the experiment. Age range varied from 24 to 46 years old. Among 10 individuals, 2 of them were males and remaining 80% were females. Regarding ethnicity, all observers came from wide range of origins such as Central Asian, from Middle East, Caucasian, Indian, European, Taiwanese, and Thai.

The order and position of the samples were random. Distance from samples to the observer was around 37 cm (Figure 1). Instructions for the experiment were given before starting the procedure. The instructions were as follows: "Please choose the PANTONE Skin Tone Guide sample colour that best matches with your skin colour. Feel free to choose as you are choosing your foundation colour at the makeup store".

Figure 1. *An observer performing our experiment. Permission of observer to publish this image was obtained*

The experiment consisted of two phases. The first one was the skin colour matching and the second one was the observers' skin measurements by the spectroradiometer. Three trials were performed under each of the two illuminations: D65 and F. The observers took around two to three minutes to adapt to the illumination before starting the task. Before each trial, samples were shuffled. It is important to mention that different observers chose different parts of their hands and/or arms in order to match their skin colour with the given samples. The observers' chosen samples under the two different illuminations were annotated for each of the three trials. Then, the reflectance of the skin part used by the observer for the selection was measured with the spectroradiometer. Moreover, the reference white patch (SphereOptics Zenith Lite of 95% reflectance) [15] radiance values for each observer's response under two illuminants and in all three trials were also captured by the spectroradiometer. This is necessary for calculating reflectance values for subsequent data analysis. The reflectance values were computed by dividing the spectral radiance of the sample by the radiance of the white, and then multiplying by the calibrated reflectance of the reference white.

Also, RMSE and GFC mean values were computed by comparing the reflectance measured under the two different illuminations for the set of all reflectance curves, with values of 0.0192 and 0.9990 respectively. These values indicate that the spectral reflectance curves measured under the two different illuminations are indeed very similar. In general, there was not any time limitation for observers to perform the matching procedure. However, 90% of observers spent around 25-30 minutes for the whole experiment including hand and arm radiance values measurements with the spectroradiometer.

Results and Discussion

Because of noise in the beginning part of spectrum and near infra-red regions, we truncated our samples in the spectral range of 400-780 nm with 2 nm interval. We analysed four different factors using the results obtained in the experiments.

Factor 1: PANTONE Skin Tone Guide

The PANTONE Skin Tone Guide samples were analysed in order to check the possibility of selecting a limited number of skin tone samples in a psychophysical experiment. In the PANTONE Skin Tone Guide, each shade is shown by a letter indicating its hue type (e.g. R for red and Y for yellow), surrounded by two numbers on the left and right showing its hue undertone and lightness levels respectively. There are five levels for hue undertone and fifteen levels for lightness.

The spectral reflectance of the PANTONE Skin Tone Guide samples was converted into XYZ tristimulus values. Then CIELAB values were calculated from XYZ coordinates. The noticeable point in this analysis is that lightness, L* is not increasing equally for all the rows of the PANTONE Skin Tone Guide samples and the change in L* values is more when lightness number decreases (i.e. L* varies more for darker shades). Moreover, CIEDE2000 colour difference values are calculated for shades with the same hues and the same lightness respectively. The results show that colour difference between rows of the PANTONE Skin Tone Guide samples increases by a decrease in the lightness while colour difference between shade samples with the same lightness, but different undertones is kept almost the same.

Due to the high values of CIEDE2000 for neighbour shades which are mostly beyond just noticeable difference (JND) that is considered $\Delta E = 0.5$ in [16], these analyses led to not subsample the samples in order to keep the precision of study to the differences between neighbouring samples.

Factor 2: Suitability of the PANTONE Skin Tone Guide sample set for representing the observers' skin colour.

We analysed if the PANTONE Skin Tone Guide samples cover all the participants' skin colour in the experiment. Therefore, XYZ tristimulus values were calculated from spectral radiances of both the PANTONE Skin Tone Guide samples and the participants' skin for CIE 2° standard observer. Subsequently, CIELAB and CIELCh values were derived and plotted. As shown in Figures 2 and 3, almost all the participants' skin colours are covered by the PANTONE Skin Tone Guide samples except for two of them which are relatively close to the samples' gamut in colour space.

Factor 3: Chosen sample by individuals versus the closest sample among the PANTONE Skin Tone Guide samples.

In this part, the efficiency of the observer's human visual system in choosing the closest skin coloured sample to their skin was studied. The average of the three trials was considered as the selected colour by the observer. Euclidean distance (i.e. CIELAB colour difference) was calculated between measured skin colour to selected sample and to the closest sample among the PANTONE Skin Tone Guide samples.

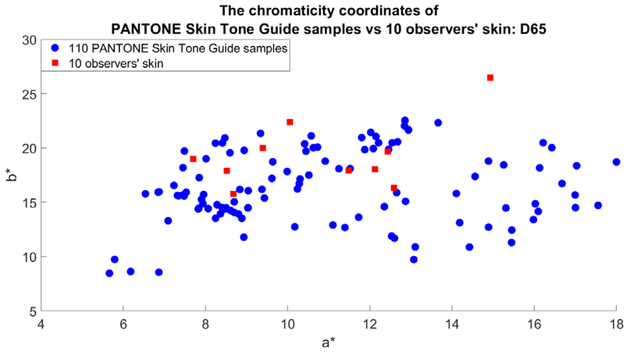

Figure 2. *The chromaticity coordinates of the PANTONE Skin Tone Guide samples (blue) and participants' skin colours (red) under D65 illuminant*

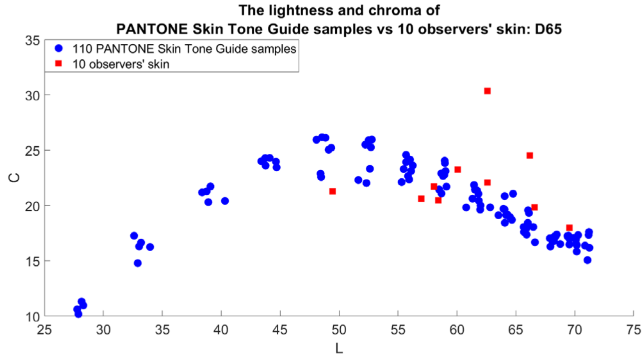

Figure 3. *The colour coordinates of the PANTONE Skin Tone Guide samples (blue) and participants' skin colours (red) under D65 illuminant*

The measurements and calculations were done separately for D65 and F standard illuminations. The results are presented in Table 1 and an illustration of these distances under D65 are presented in Figures 4 and 5.

Table 1. CIEDE2000 colour differences between actual skin colour, selected sample, and the closest sample

	Colour difference under D65		
Observer	Skin to selected	Skin to the closest	Selected to the closest
Mean	7.59	2.50	6.31
Std	4.43	2.18	4.83
	Colour difference under F		
Observer	Skin to selected	Skin to the closest	Selected to the closest
Mean	8.75	2.81	8.04
Std	4.44	2.22	4.85

From this table, it can be deduced that selection by observers was slightly better under D65 than under F illuminant because the colour difference between selected to the closest samples is smaller for D65. Also, the difference between skin and selected sample is lower under D65.

It is noticeable that distance between the selected sample by observer to the measured skin is about three times more than the distance between the closest sample and measured skin, and it means that instrument could find the best match much better the human visual system. We attempt to explain this relatively poor performance of the human visual system of observer versus objective instrument evaluations with the fact that each

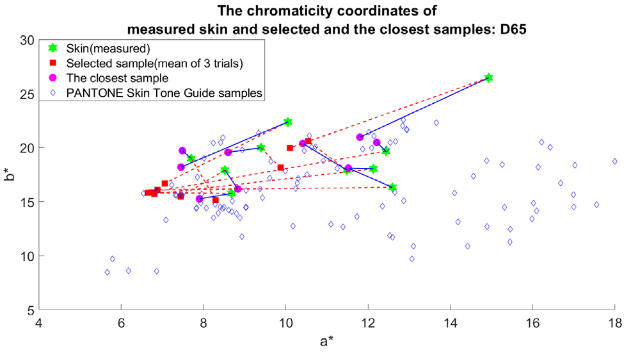

***Figure 4.** The chromaticity coordinates of actual skin, selected sample, and the closest sample under D65 illuminant*

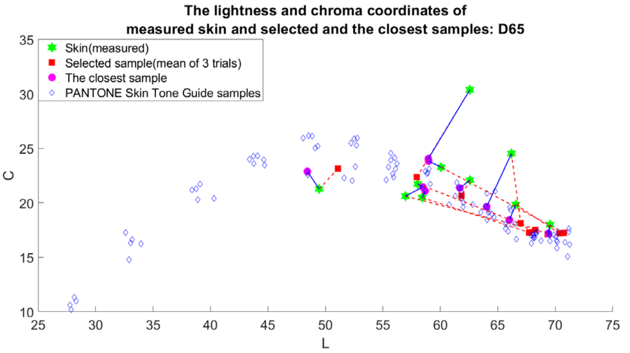

***Figure 5.** The colour coordinates of actual skin, selected sample, and the closest sample under D65 illuminant*

observer is unique, and their responses can vary from each other's.

It is obvious that observers tend to choose less reddish and consequently paler samples in comparison to their skin colour (Figure 4). From Figure 5, it is obvious that observers tend to choose lighter samples and with less chrominance samples in comparison to their skins. This trend is the same for both illuminations.

Factor 4: Intra and inter observer variability.

The majority of observers performed better under D65 than under F illuminant (Table 2). Also, overall mean values reveal that under D65 (mean MCDM value is 0.9047 and standard deviation 0.4270), people tend to select colours more consistently rather than under F illuminant (mean value is 0.9238 and standard deviation 0.6684). Interestingly, our observers told that colours under F illuminant seemed for them closer to skin colours. Both minimum and maximum values of MCDM for D65 illuminant were smaller in comparison with F illuminant.

Table 2. Intra-observer variability

	MCDM D65	MCDM F
Mean	0.9047	0.9238
Std	0.4270	0.6684
Min	1.2616	1.3006
Max	2.2705	2.8313
Median	1.6575	1.4843
95^{th} percentile	2.2596	2.7150
5^{th} percentile	1.3375	1.3148

Table 3 shows comparison between inter and intra observer variability. As it can be seen, mean value for inter-observer variability is higher than mean value for intra-observer variability under both illuminants. Under F illuminant mean values for both intra and inter observer variability are slightly greater than under D65. We assume this significant difference of inter-observer variability might be due to the inherent variability in the observers' complexions and skin colours.

Table 3. Inter-observer variability

	MCDM D65	MCDM F
Mean intra	0.9047	0.9238
Mean inter	5.5074	5.3347

Conclusions and future work

In this study, the efficiency of human visual system when choosing the best foundation colour match was evaluated through a psychophysical experiment. 110 skin colour samples of the PANTONE Skin Tone Guide were shown to the observers and they were asked to select the closest sample to their skin tone. The procedure was done under D65 and F illumination and repeated three times under each light source to investigate repeatability of the experiment. Both the PANTONE Skin Tone Guide samples as well as the participants' skin were measured using a spectroradiometer. All the reflectance factors were calculated and analysed in terms of repeatability of observer and instrument. In addition, CIELAB values were computed and the colour differences between observers' selected colour and instrument measurements were compared.

The results showed that selected PANTONE Skin Tone Guide samples are adequate for our psychophysical experiment as they cover our observers' data. Also, inter and intra observer variability via MCDM are computed and revealed that, in terms of intra-observer variability, our observers were consistent with their decision about colour match under each trial and each illuminant.

Furthermore, skin colour assessment was done better under D65 than under F illuminant. The human visual system was three times worse than instrument in colour matching in terms of colour difference values and observers tend to choose lighter, less reddish, and consequently paler colours than their own skin colour.

As future works, the experiment will be done with more participants and texture of the human skin will be taken into account. Moreover, consideration of skin colour shades samples made of materials closer to the real human skin will definitely improve the results. Also, we will extend the study with light emitting diode (LED) as LED is becoming popular as indoor light source in beauty stores.

References

[1] J. Jain, N. Bhatti, H. Baker, H. Chao, M. Dekhil, M. Harville, N. Lyons, J. Schettino, and S. Susstrunk, "Colour match: an imaging based mobile cosmetics advisory service." In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, pp. 331-334. ACM (2008).

[2] B. Jones, 2006, "Approximating the Appearance of Human Skin in Computer Graphics."

[3] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods." Pattern recognition 40, no. 3 (2007): 1106-1122.

[4] M. Störring, "Computer vision and human skin colour." Doctoral dissertation, Computer Vision & Media Technology Laboratory, Faculty of Engineering and Science, Aalborg University.

[5] J. Welzel, S. Schuh, “Noninvasive diagnosis in dermatology.” JDDG: Journal der Deutschen Dermatologischen Gesellschaft. 2017 Oct;15(10):999-1016.

[6] P. Burger, A. Landreau, S. Azoulay, T. Michel, and X. Fernandez, "Skin whitening cosmetics: Feedback and challenges in the development of natural skin lighteners." Cosmetics 3, no. 4 (2016): 36.

[7] P. J. Matts, "New insights into skin appearance and measurement." In Journal of Investigative Dermatology Symposium Proceedings, vol. 13, no. 1, pp. 6-9. Elsevier, 2008.

[8] T. Hirotsugu, "Measurement of skin color: practical application and theoretical considerations." Journal of Medical Investigation 44 (1998): 121-126.

[9] I. L. Weatherall, and B. D. Coombs, "Skin color measurements in terms of CIELAB color space values." Journal of investigative dermatology 99, no. 4 (1992): 468-473.

[10] I. M. Gibson, "Measurement of skin colour in-vivo." J Soc Cosmet Chem 22 (1971): 725-40.

[11] A. Treesirichod, S. Chansakulporn, P. Wattanapan, “Correlation between skin color evaluation by skin color scale chart and narrowband reflectance spectrophotometer.” Indian journal of dermatology. 2014 Jul;59(4):339.

[12] W. Yuzhao, M. R. Luo, M. Wang, K. Xiao, and M. Pointer, "Spectrophotometric measurement of human skin colour." Colour Research & Application 42, no. 6 (2017): 764-774.

[13] J. Romero, A. Garcıa-Beltrán, and J. Hernández-Andrés, "Linear bases for representation of natural and artificial illuminants." JOSA A 14, no. 5 (1997): 1007-1014.

[14] R. S. Berns, “Billmeyer and Saltzman’s principles of color technology.” Willy, New York, USA, 2000, pg. 97.

[15] S. D. Noble, A. Boeré, T. Kondratowicz, T. G. Crowe, R. B. Brown, D. A. Naylor, “Characterization of a low-cost diffuse reflectance coating.” Canadian Journal of Remote Sensing. 2008 Jan 1;34(2):68-76.

[16] Y. Yang, J. Ming, and N. Yu, "Color image quality assessment based on CIEDE2000." Advances in Multimedia 2012 (2012): 11.