A pilot study on evaluating common appearance and a colour naming approach to measure it

Philipp Tröster, Robin Schwanse, Andreas Kraushaar; Fogra institute; Munich, Germany

Abstract

If an image, for example a company logo is shown on different devices the degree of colour consistency amongst this set of stimuli can be defined as common appearance (CA). Because of the expanding number of devices with high dynamic range of imaging and with highly varying colour gamuts this issue becomes more and more important. Unfortunately conventional and established colourimetry, where usually two patches or images are compared side by side or through a media relative approach, is not sufficient, because the used metrics are developed for small colour distances. Establishing a measurement-tool for large colour differences and multiple reproductions side by side requires approaches that go beyond classic colourimetry. In this paper we designed a psychophysical experiment to measure if CA can be measured and we show a first approach for establishing a metric for CA through colour naming.

Introduction

The conventional colour reproduction methods, which are widely spread throughout the graphic arts industry, are usually based on and limited to the reproduction of the three integrated parameters *X*, *Y* and *Z*, which were defined by the international commission in illumination (CIE)[1]. Based on these parameters and the related CIELab values (ISO 11664-4), the conventional colour metrics were defined, which are the foundation of most of today's print reproduction processes. The reproduction requirements of print jobs on different printers are typically tested through a side-by-side comparison, where the original (e.g. a contract proof or a former production copy) and the reproduction are assessed under the same conditions, i. e. by the same observer and under the same norm lighting. In conventional offset printing, where different printing conditions are characterized by colour gamuts that are similar in size and shape, this approach is feasible and well-established.

In digital printing, however, the variety of gamut shapes and sizes is considerably higher, because of the wide variety of techniques and materials used. This variety of gamut shapes becomes even more pronounced, if not only hard copies but also other colour reproduction media, like digital signage monitors, come into play. These use additive colour mixing and compared with print referred reproductions result in a huge variety of colour gamut shapes and sizes. As a service provider in the graphic arts industry one aims to harvest the maximal potential of all these different gamuts when, for example, printing pictorial imagery for a stand at a fair trade on various media. This constraint prohibits the simple remedy of strictly matching a minimum colour difference between the different reproductions on the gamut intersections of the media.

Since conventional colour metrics cannot reasonably be applied, when one wants to exploit the full potential of each gamut, a somewhat looser relationship with the reference is required. Still one does not simply desire to exploit the full colour gamut, but rather achieve a 'common-appearance' between the different instances of a motive. As a good example serves Fig. 1 from Ján Morovič, which describes three approaches to display an image on three exemplary printers (A, B, C) [1]. Clearly one can realize, that both the use of a minimum colour difference ΔE in gamut intersection and the use of the 'native' colour (printing as-is in the so called device mode), without any clever gamut mapping strategy, lead to poor results. The last row of Fig. 1, on the contrary, depicts a pleasing common appearance.

***Figure 1**. Three strategies to display a reference image (left side) on different printers. The native colour approach uses the CMYK data of the reference without applying any colour management, the colour match approach uses conventional colour reproduction, and the consistent colour appearance uses an intelligent gamut mapping strategy [1].*

The consistent colour appearance of the last row of Fig. 1 is a result of a clever gamut-mapping strategy. For an optimally pleasing appearance of an image after being mapped into another colour gamut, the gamut mapping strategy cannot only consider each point of the colour space individually, but rather focus on the relationship of the colours amongst each other. The high quantity of different motives makes the search for a universally valid mapping-strategy so difficult. As Morovič states in [2] gamut-mapping is mostly a mixture of science and art, depending on the specific task, which one wants to solve. The question is, if there are certain rules, instructions or standards, which could be

---

[1] www.cie.co.at

universally applied in order to distinguish gamut mapping strategies yielding a ‘good’ common appearance from the ones which provide unpleasant results.

In order to tackle this question, we use the concept of colour naming or colour categorization, which is a much more suitable tool for the large colour differences we are dealing with. In psychology, linguistic and philosophy colour naming has a vast history and there are numerous publications, of which a few are listed in the bibliography [3-7].

Experimental setup

The following psychophysical experiment focuses on the detection of common appearance, defined as the degree of visual consistency among a set of stimuli. Or in other words, does the average observer of a several sets of stimuli always prefer a particular set, because it seems more pleasing than the others. The simple concept of the experiment is explained in Fig. 2.

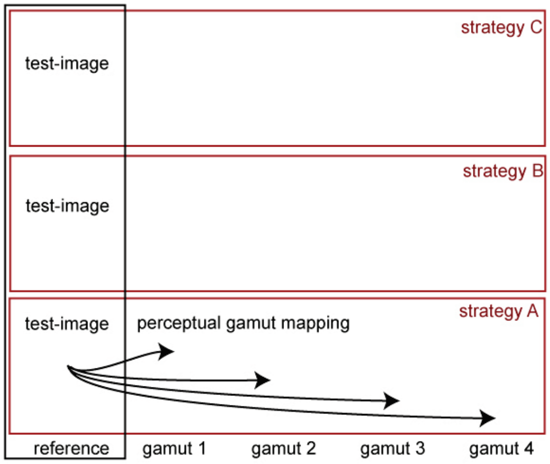

***Figure 2**. The concept of the experimental setup: A test-image, originating from a reference gamut is mapped to several gamuts via different perceptual mapping strategies.*

If someone wants to present an image such as a company logo on several media in a way, that they “fit to each other” usually the perceptual gamut-mapping approach is chosen to map the data. This mapping approach is, in contrast to the absolute and relative colourimetric rendering intents (RI), not mathematically defined. Thus the result depends on the profile maker and is hence “50% art and 50% science” [2]. Therefore each software package, which offers this RI does it differently and claims that their approach results in common appearance. The gamut-mapping strategies used in this experiment are listed in Table 1.

Table1: Software packages performing the perceptual gamut-mapping. Strategy C represents the native approach (Fig. 1).

A	X-Rite i1 profiler (v.1.5.6)
B	Heidelberg Colour Tool (v.13.00.31.7)
C	Device Mode (no gamut mapping)
D	basICColour print 3 (v.3.1.0)
E	Color Logic CoPrA 3 (v.3.2.2)
F	Agfa ColourTune (v.8.0.1.8)

In this study the colour-gamuts used for this experiment are CMYK-based and represent typical printing conditions of the graphic arts industry. As the reference the gamut of the Epson Stylus Pro 9900 Proofing-System with semi-matte paper is used. The further gamuts are Fogra39, CRPC 4, CRPC 1 and the Inkjet-gamut. CRPC1 and CRPC4 are members of the Characterized Reference Printing Condition of the ISO/FDIS 15339-2 standard and represent general printing on super-calendared paper and typical newsprint. Fogra 39 is typically used in commercial printing. The inkjet-gamut, however, was developed to represent a typical inkjet-gamut and to serve as an exchange-space for digital inkjet-printing. All used colour gamuts are shown in Fig. 3. The reference gamut comprises the other gamuts, which enables the absolute colourimetric mapping of the test-image back to the reference gamut after the perceptive mapping. Ensures, that each motif can be printed on the Epson proofing system after the mapping is done.

As can be recognized is their shape and form varying to an extent, that an absolute colourimetric match of an image would lead to severe restrictions. Still the variation of the used gamuts is not excessively high considering the different gamut-shapes in digital printing, or if one takes also RGB-base gamuts into account. Therefore if CA can be measured applying the used gamuts (a-d) one expects an even stronger effect using further gamuts.

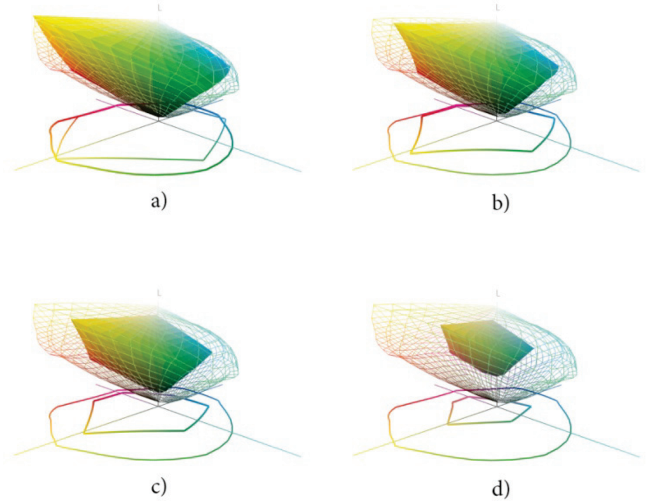

***Figure 3**. The colour gamuts used in this study; the wireframe represents the reference gamut, from which the test-image is mapped to a) Inkjet-gamut b) Fogra39 c) CRPC4 and d) CRPC1.*

The degree to which the quality of a certain gamut-mapping strategy and therefore its CA is dependent on the considered test-image is unknown. Of course the software-developers themselves carry out an intense amount of tests with many test-images in order to optimize their product. In the context of this experiment, which is part of Fogra research project 10.058 [8], the aim is get many observers in a manageable amount of time. Therefore one carefully chosen test-image is used. The image contains the colour-circle of the ISO 12647-7 evaluation test-chart and a grayscale separating the grey-axes into 10 equidistant parts. This way the treatment of both saturated and non-saturated colours is taken into account.

Additionally the neighborly relationships between colours are taken into account. The composition of the test-image of one of the strategies in Table 1 is shown in Fig. 4.The colour circle and the grayscale is mapped from the reference to the various gamuts, resulting in 120 colour-fields and 50 greyscale-fields.

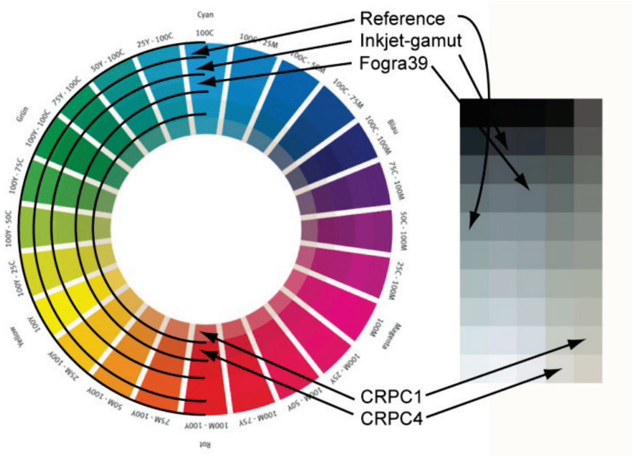

***Figure 4**. Test-image used for our psychophysical experiment exemplarily for one gamut-mapping strategy (A-F).The 24 colour-fields and the 10 grey-scale-fields are mapped from the reference gamut to the four gamuts.*

The study consists of two experiments, both with the same set of test-images created through the strategies A-F. This is shown in Fig. 5. The first is a rank order experiment in which 38 participants, who are all normal sighted, participated. Half of these persons are experienced with colour-matching. The second experiment is a pair-comparison experiment with forced choice in which was carried out by 15 participants. In both experiments instructions were given in written form. The spectators of the test-image were asked to judge the consistency of the colours without further specifying if they should focus more on the colour-circle or the grey-scale. Some of the participants were stating, that they would rank the strategies different, if they would judge only the colour-circle or the grey-scale. Those persons were told, that they should consider both and decide for themselves which is more important for them.

***Figure 5**. Setup of the rank-order experiment (left) and the pair-comparison experiment (right). Both are carried out under D50 norm-light.The instructions were given in written form.*

Results of the rank order experiment

The distributions of the rankings of the stimuli A-F, their mean opinion score and the according standard deviation are shown in Fig. 6.

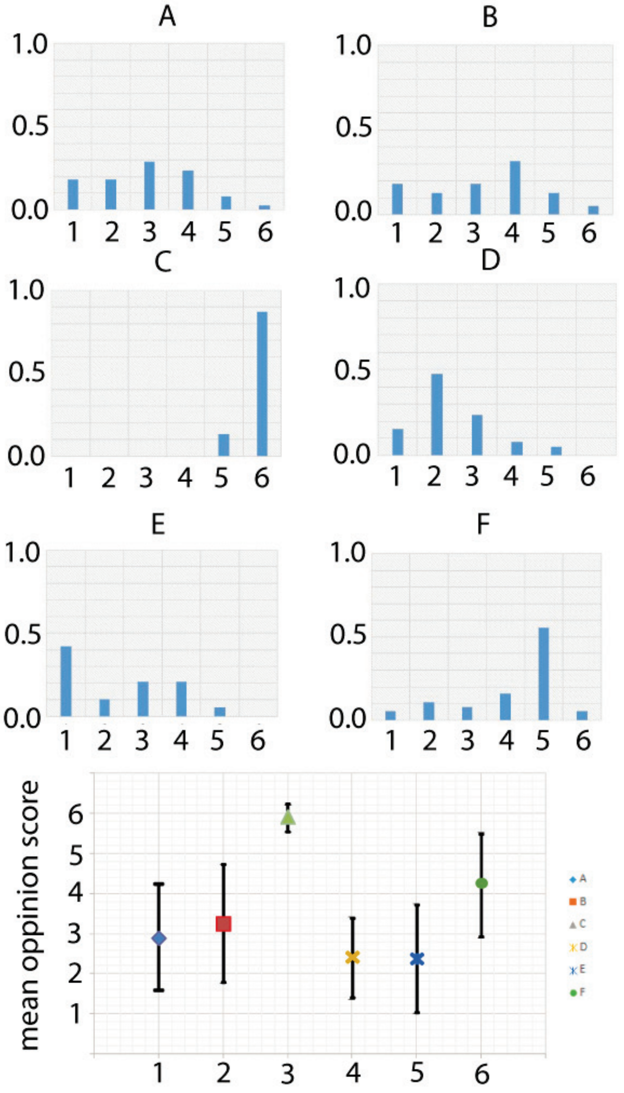

***Figure 6**. Setup of the rank-order experiment (left) and the pair-comparison experiment (right). Both are carried out under D50 norm-light.The instructions were given in written form.*

As expected the device-mode approach, or native approach, is clearly judged as providing the worst consistency amongst the stimuli. Therefore a unimodal distribution can be seen for stimulus C. Also stimulus D and F show clear unimodal distributions ranking them second and fifth. For the three remaining stimuli A, B and E, the distribution is still smeared over a wide range and it is obvious, that more statistics is needed in order to judge the results. This is also reflected by the mean opinion score that results from this data and the according standard deviation. Several participants were also stating that a judgment is hard, because there are too many stimuli at the same time. Therefore the pair comparison experiment was carried out.

Results of the pair comparison experiment

The result of the pair comparison experiment with forced choice is presented as scores applying Thurstone's model Case V [18] and the according 95% confidence interval. This data is shown in Fig. 7.

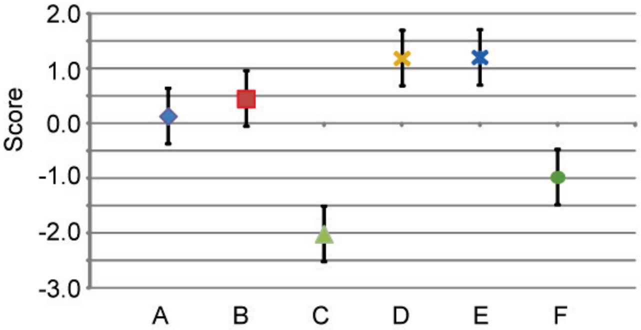

***Figure 7**. The results of the pair comparison experiment shown as the score of Thurstone's Case V model. Strategy D and E are statistically indistinguishable as well as the strategies A and B.*

The data of this experiment not only show a correlation with the mean opinion score of Fig. 6, but also indicate, that using a pair wise comparison the result converges more quickly than in the rank-order experiment. Both data sets indicate, that stimuli D and E are indistinguishable regarding their CA. The same is valid for stimuli A and B. These two groups can be considered as showing the best and the second best degree of colour consistency, followed by mapping strategy F and the native approach C.

Connecting CA to Colour Naming

As briefly stated in the introduction the conventional colourimetry is not able to deal with such large colour differences as present in this study. Therefore colourimetry does not seem to be a feasible tool to provide a metric for CA. the field of Colour-Naming, however, seems to be more promising. Colour-Naming is closely related to linguistic relativity, which was made popular a study of Berlin and Kay [7] and which deals with the connection of language and thought. There is a debate going on if colour-terminology has absolute universal constraints or if it is dependent on cultural background and language. In order to examine this issue several monitor-based online-experiments were carried out trying to find out the names of colours, their frequency and their exact position in the colour space.

Two of these studies are the colour naming experiment of Giordano Beretta and Nathan Moroney [9, 10] and the Colour Naming Experiment of Dimitris Mylonas [17]. Both monitor-based experiments show the participants several colour patches, asking the observers to name them. The used sRGB values of the patches are statistically evaluated connected with a colour-name. The study of Moroney results in 746 names and the Study of Mylonas in 489. The sRGB triplets can be converted to CIELab values and are assumed to be the centers of volumes in the colour space, which belong to a certain colour name. Basically one ends up with a heterogeneous colour-name-density in the Lab-space.

In order to connect the findings shown in Fig. 7 with the results of the colour-naming experiments, we assume, that the more colour-names are "overstepped" when mapping the Lab-values of the colour-circle and the grey-scale of the reference to the other gamuts, the more it disturbs the spectators impression of common-appearance. Exaggeratedly spoken: "If red turns green, it would bother." Therefore as a first simple approach we intent to count the amount of colour-names, that lie between the Lab-point of the colour-fields of the reference and the according Lab-points of the various instances. To decide if a colour name is crossed or not, we simply connect two Lab-points through a linear slope. Then each colour-name is situated between them if its distance to the connecting line between the two Lab points is closer than a threshold. The principle is shown in Fig. 8.

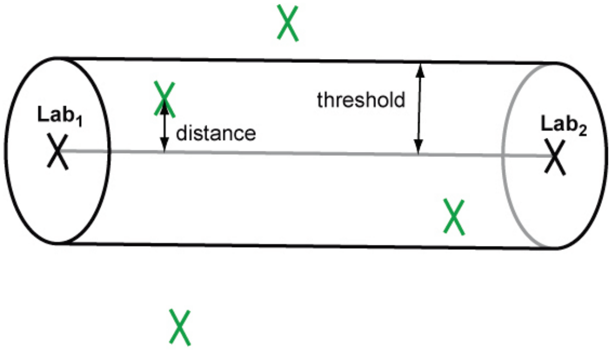

***Figure 8**. Principle of counting the amount of colour-names (green crosses) situated in between the Lab-point (Lab1) of a colour-field of a reference and one of the fields of the other gamuts (Lab2). The color names, which are closer to the connection of Lab1 and Lab2 than the threshold are counted as crossed colour names. In this example there are two colour-names in between them.*

The question is how to choose the threshold that determines the shape of the cylinder and, therefore, how many color names are crossed. In this study it was chosen as a function of the distance between the two Lab-values:

$$threshold(d) = s \cdot (1 - \exp[-k \cdot d]).$$

The parameter s represents the limit for larges distances between Lab_1 and Lab_2. For small distances the parameter k kicks in, which makes sure that they do not contribute over proportionally to the total amount of colour-names. The values of the two parameters $s = 3$ and $k = 0.1$ can be considered as an educated guess and must be examined further.

To characterize a whole strategy by colour-names we summarize the crossed colour-names over all combinations of colour-spaces (cs) and all colour fields (cf) of the colour-circle and the grey-scale. To illustrate this principle the positions of the colour-names of both studies [9,10,17] projected to the ab-plane can be seen in Fig. 9. Additionally the ab-values of the colour-fields of the test-image in the different gamuts, mapped by strategy A, are shown. The Lab_1-value and the Lab2-value of Fig. 8

correspond to the coloured crosses of Fig. 9 while the green crosses of Fig. 8 represent a selection of the grey ones in Fig. 9.

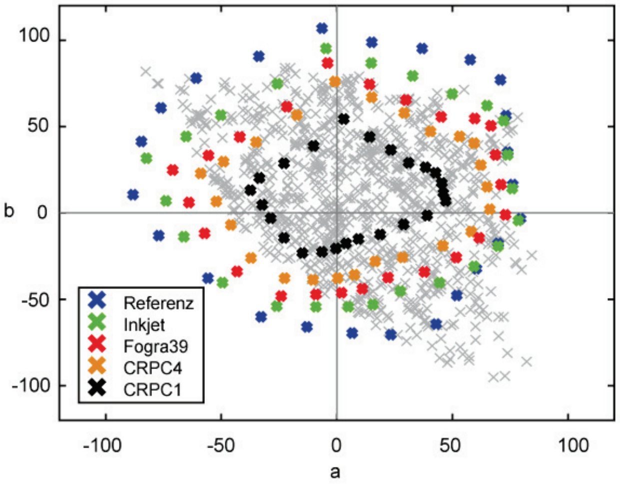

***Figure 9**. The positions of the colour-names of the two experiments [9,10,17] projected to the ab-plane are plotted as grey crosses. The coloured crosses give the positions of the colour-fields of the test-image of strategy A.*

One can recognize the characteristic shape of the sRGB-gamut, which does not expand to the cyan region in the third quadrant of the ab-plane, contrary to the typical CMYK-based gamuts used in this study. Therefore we scale the positions of the colour-names. In order to do this the colour-names are sorted to 24 bins, depending on their hue. In each of these angle-dependent bins the chroma C_{ref} of the correspondent Lab-value of the colour-fields of the reference-gamut (cf. blue crosses in Fig. 9) and the maximal chroma $C_{cn,max}$ of the colour names in this bin are detected. Then each colour-name position in each bin is scaled with the corresponding factor $f = C_{ref} / C_{cn,max}$. After carrying out these steps one ends up with colour-names, which fit to the shape of the reference colour-gamut, as can be seen in Fig. 10.

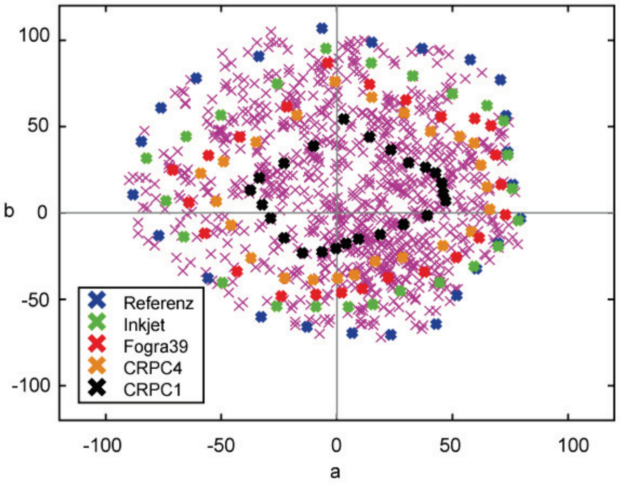

***Figure 10**. The reshaped positions of the colour-names of the two experiments [9,10,17] projected to the ab-plane. The colour-name density is obviously not uniform.*

Please note that the lightness is not ignored in this study even though the graphs only show the projection to the Lab plane. In order to carry out the operations leading from the original colour-name positions the reshaped ones, it is assumed, that the position of the colour-names can be interpreted as media-relative. That means, that for example the name of the brightest green in one colour-space will always be bright-green, but its position in the Lab-space will differ according to which medium was used during the evaluation of the colour-names. Of course this assumption has to be investigated further in future experiments.

After reshaping both colour-name spaces from the two experiments, which we combined for our calculations, the number of overstepped colour-names is determined. The result is plotted against the scores of Thurstone's Model Case V shown in Fig. 7.

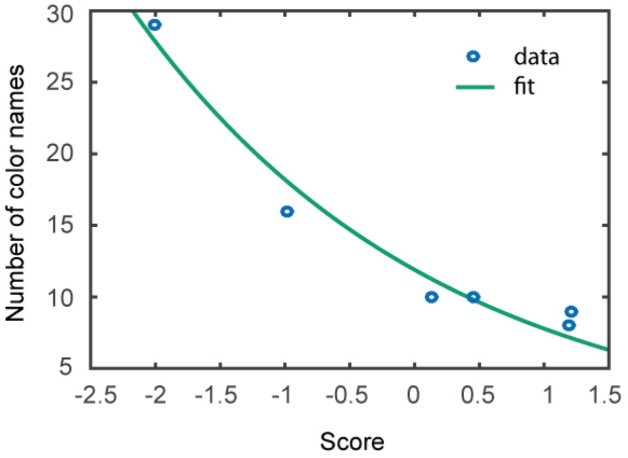

***Figure 11**. The number of crossed colour-names (explained in figure 8) of the different strategies (A-E) are plotted against the Score of Thurstones Model-V (blue circles). The green curve is an exponential fit of the data. The data shows a promising correlation.*

The blue circles show, that the better the CA appearance of a stimulus is judged, the less colour names are crossed. The indistinguishability within the two groups (A,B) and (D,E) can be seen also in this plot. The data is fitted to an exponential function, which is shown as a green line.

Discussion and Conclusion

The scientific evaluation of CA, as defined in the introduction, is still a brand new topic. As a first approach to get to the bottom of the questions "is there something called CA?" and if answered yes "how can we measure it?" is done by this experiment. As is indicated in both the results of the rank-order experiment and the pair-comparison experiment, CA is measurable. However, more statistical data is needed to further strengthen this statement.

The approach to measure CA by using a set of colour-names shows also promising first results but needs to be examined way further. First of all it needs to be established to what extent the data sets given in the colour-naming experiments are given are valid for the experiment that is done here. Second it needs to be determined whether the conducted manipulation of the positions of the colour-names is valid. Third the way in which the colour names are counted has to be examined further. Especially the relationship of neighboring colour fields, and how they change throughout a gamut-mapping should be included into a possible metric of CA.

References

[1] Morovič J.: *Colour references for digital print production*. (2010).

[2] Morovič J.: *Colour gamut mapping*. John Wiley & Sons. USA, (2008).

[3] Kaufmann R., O'Neill M. C..: *Colour names and focal colours on electronic displays*. Ergonomics, pp. 881-890, (1993).

[4] McFadden S., Kaufmann R., Janzen M.: *Classification of colours presented against different coloured backgrounds.* Displays, **15**, pp. 203-214, (1994).

[5] Regier T., Kay P., Khetarpal N.: *Colour naming reflects optimal partitions of colour space*. Proc. Natl. Acad. Sci. USA, pp. 1436-1441, (2007).

[6] Schirillo J. A.: *Tutorial on the importance of colour in language and culture.* Colour Research & Application, **26**, pp. 179-192, (2001).

[7] Berlin B., Kay P.: *Basic Colour Terms: Their Universality and Evolution*. CSLI Publications, Stanford, California, (1999)

[8] *Einflussgrößen auf die hochqualitative Monitordarstellung hinsichtlich einer Standardisierung von Softproofsystemen*. München: Fogra (2014) (Nr. 10.058) – Forschungsantrag

[9] Beretta G. B., Moroney B. M.; Hewlett-Packard Laboratories, pp. 1-27, (2010).

[10] Moroney N.; Electronic Imaging pp. 36-46, (2003).

[11] Fairchild M.: *Colour Appearance Models*. Wiley & Sons, West Sussex - England, pp. 310 ff.

[13] Engeldrum P.; Winchester: Imcotec Press, pp. 257-261, (2000).

[14] Zolliker P., Simon K.; J. Electron. Imaging, **15**, Int. Soc. For Optics and Photonics, pp. 013004, (2006).

[15] Zolliker P., Simon K.: Conf. On Colour in Graphics, Imaging and Vision, pp. 257-261, (2006).

[17] Thurstone, L. L.;A law of comparative judgment. Psychological Review, vol. 34, pp. 274-286 (1927).

[18] Mylonas, D., & MacDonald, L.; Colour in Graphics, Imaging. Joensuu, Finland: IS&T (2010).

Author Biography

Philipp Tröster received his diploma in physics from the technical university of Karlsruhe (2009) and his PhD in biophysics from the Ludwig-Maximilians Universität in Munich (2014). Then he entered the Fogra research institute in Munich where he works as Scientist and developer. Amongst other topics as 3D-printing and energy-efficiency, he is doing research in the field of applied colour-science.