Evaluation of mask experiment – conducted PA meeting 2012-11-30

Problem / Question

1) Likely due the peaky spectra of monitors observer metamerism might become more evident for softproofing then for comparing proof prints. Is this true – how big are differences between observers when doing a Softproof appraisal?
2) How good is the correlation of a laboratory spectroradiometer measurement to the visual appearance?

Experiment

To get closer to an answer for both both questions an experiment was conducted where a white paper (without OBA) was mounted on a grey neutral background. A hole of similar size of the paper (2 cm) was cutted direct besides the paper, so a 2 cm x2 cm portion of the monitor was viewed in direct comparison to the paper. The observers were asked to

a) Adjust the monitor via L*, a* and b* equivalent controls to match the paper in brightness and chromaticity and to
b) Denote how good the match achieved was judged by the observer himself visually. The observer could choose between
    a. No colour difference
    b. Small colour difference
    c. Medium colour difference
    d. Big colour difference

The colours of the illuminated white paper and the monitor was measured with a laboratory Tele-Spectroradiometer (Konica Minolta CS-2000).

Figure 1: Photo illustrating the experiment setup and the measurement. The actual experiment was conducted with the exclusive illumination of the viewing cabinet. On the photo the ceiling lights are added for better visability of the setup only.

Evaluation

Reproducibilty of the experiment

Measurements

Due practical reasons it was not possible to immediately measure the monitor and the illuminated paper. The monitor and the illuminated paper was measured a few hours after the experiment (while the monitor and viewing cabinet was still warm) and additionally three days later after a new completed warmup-phase of all components. The monitor measurements showed a variation about 1,5 thousands in CIEx and about 0,5 thousands in CIEy. These difference relate to a small colour difference of approximatly ΔE00 = 1 (see figure). Luminance change was negligible.

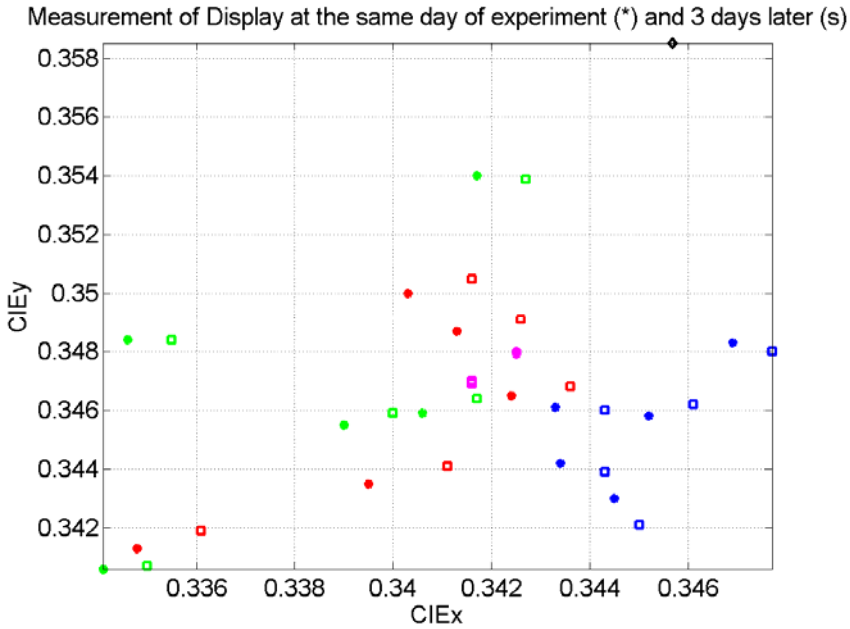

Figure 2: Sample measurement of three observers, for 5 repated judgements on day one (dot symbol) and 3 days later (square symbol).

The paper white CIEYxy measurements showed a slightly bigger change (115 / 0.3425 / 0.3479 for the first day and 112 / 0.3416 / 0.3469 three days later).

Observers

Each observer repeated the visual match 5 times. The variations of the 23 observers are seen in the following figure. The luminance match CIEY showed bigger variations then the chromaticity match CIExy.

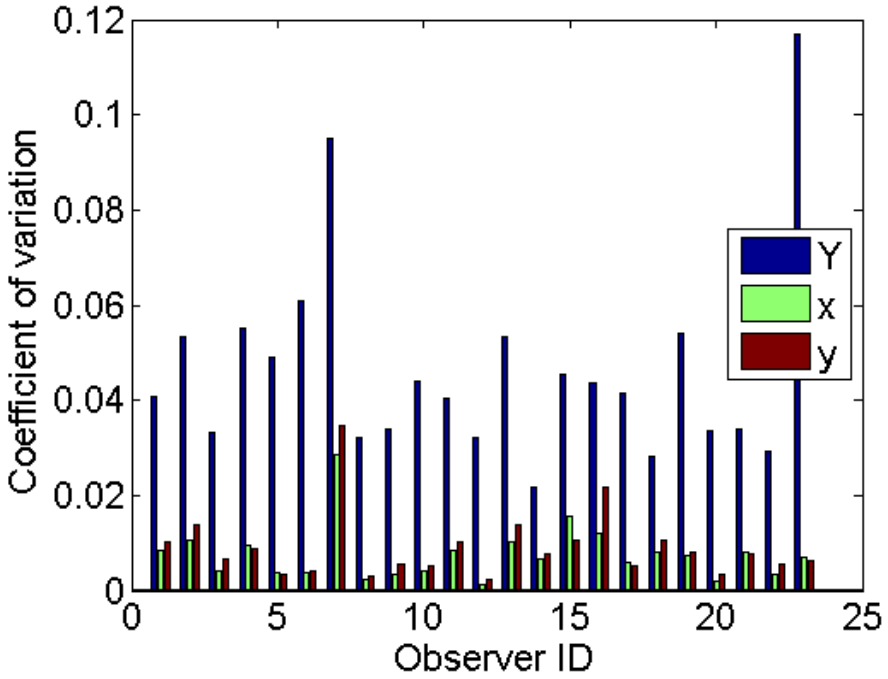

Figure 3: Variations of 5 repated visual matches of the 23 observers. Chromaticity CIExy mostly showed a variation of about 1 %.

Absolute variations are shown by means of the standard deviation for CIEY and CIExy.

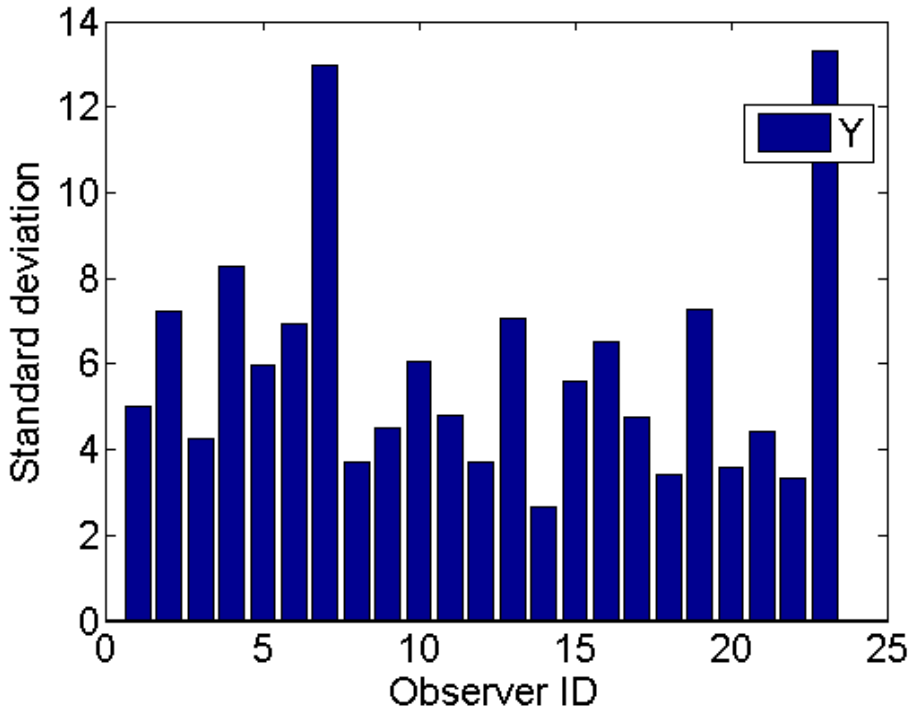

Figure 4: Average standard deviation of Y between 23 observers is 5,9 cd/m².

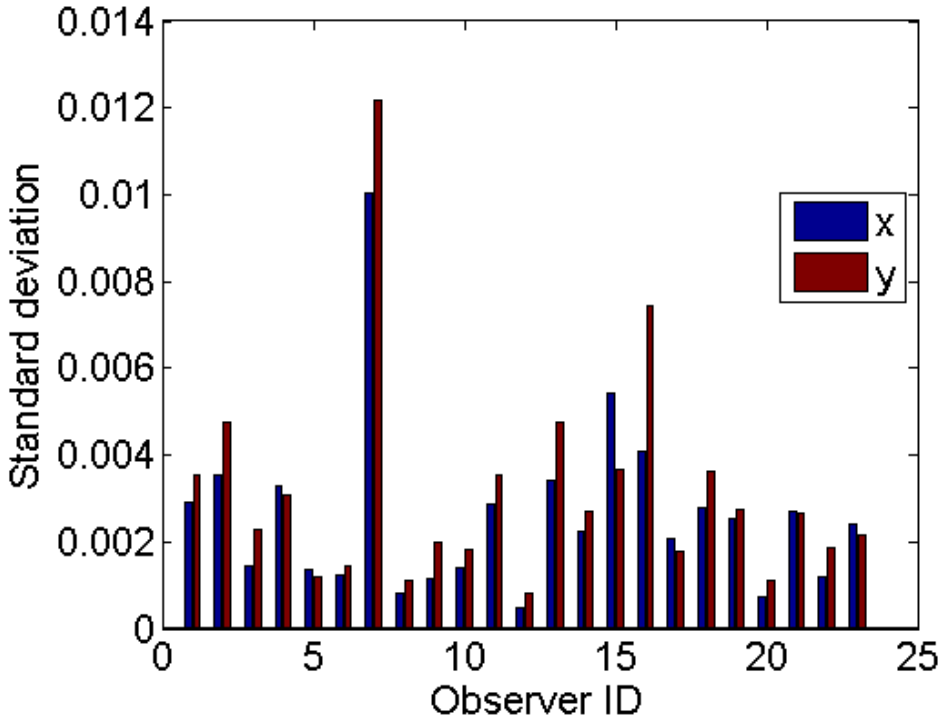

Figure 5: Average deviation between 23 observers: CIEx = 0.0027 and CIEy = 0.0031.

It can be seen that some observers are fairly consistent, while others show a much higher variation for repated colour matches. Observer 7 showed the biggest variation, but was not as satisfied as most other observes with his matches (see following figure). Also Observer 15 showes a higher than average variation and was not as satisfied as most observers. These indicates a potential correlation between variance and satisfaction with the own generated “best match”, although other observers (ID 4 and ID 21) state also a small to medium colour difference while showing average variation.

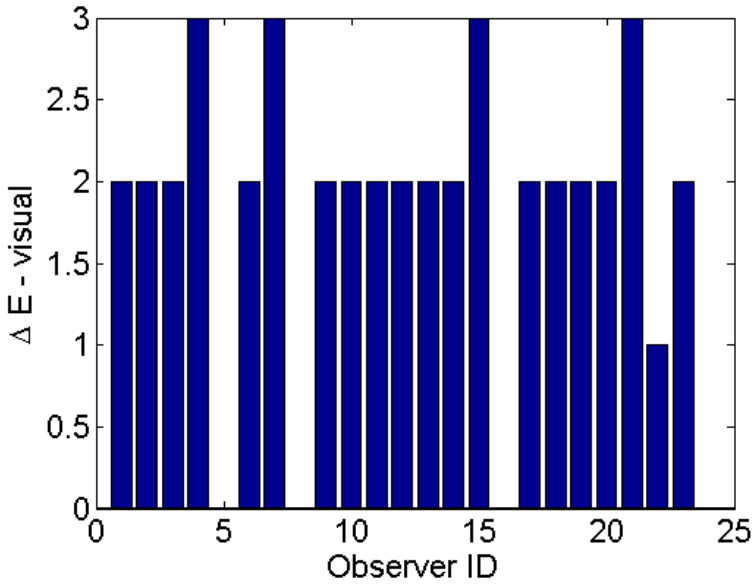

Figure 6: Judgement of each observer himself how good the achieved colour match was. No colour difference is denoted as 0, a just noticable colour difference denoted as 1, a small colour difference denoted as 2 and a small to medium colour difference denoted as 3.

Conclusion: For further experiments an anchor pair representing the four possible colour differences (“no”, “small”, “medium”, “big”) will be presented where a luminance difference of grey will exhibit a ΔY of 2, 4 and 6. Also allowing to choose an colour difference between the presented colour difference pairs this will result in 7 possible choices correlating to ΔE = 0 to ΔE00 = 7 (for colour differences higher than the pair where ΔY = 6).

Results of experiment

The average of 5 visual matches of each observer where averaged. The following figure shows the range of the visual matches in chromaticity.

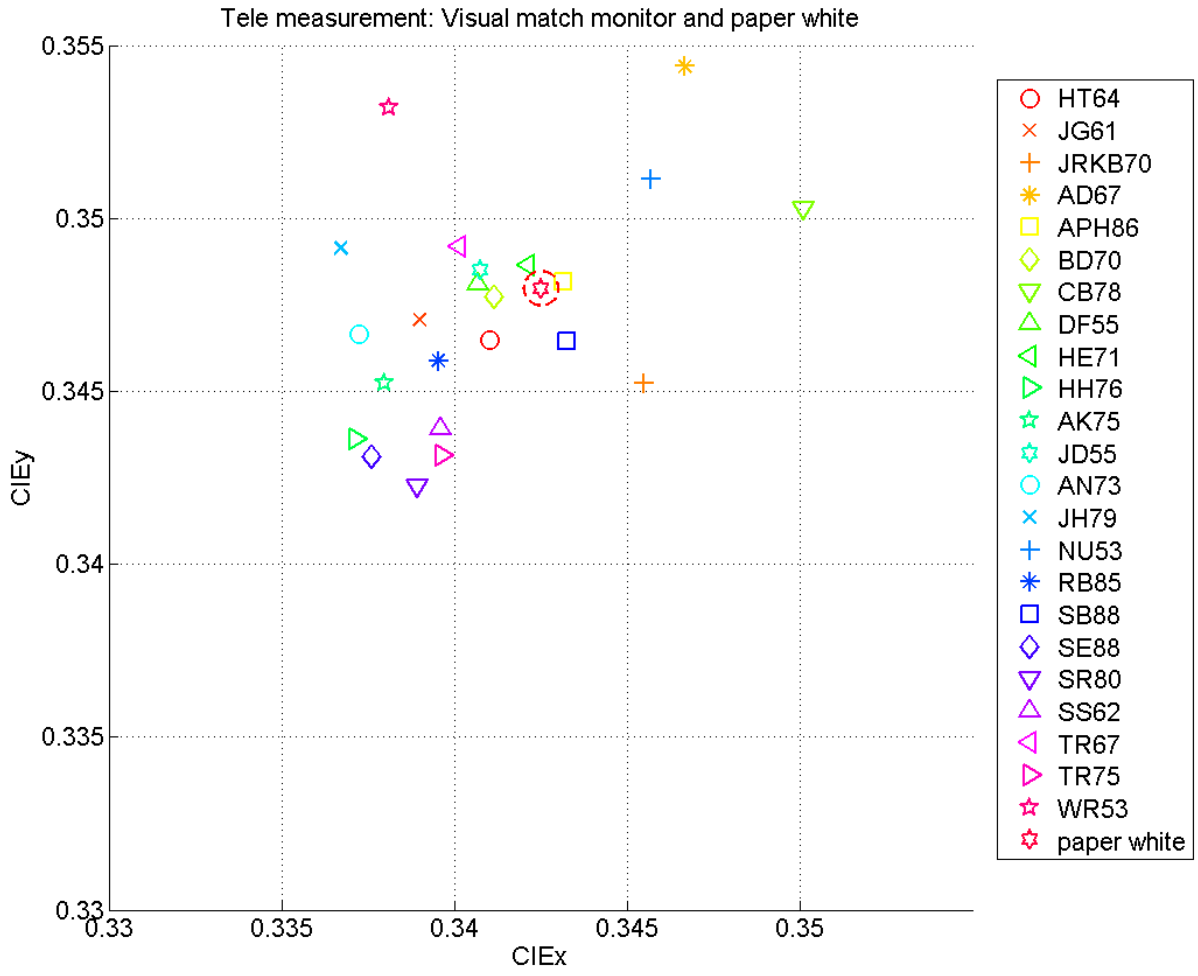

Figure 7: Average of 5 visual "best matches" between a montior and paper white for 23 observers. Each indiviudal observer is denoted with a short name, where the number part stands for the birth year. The circle of the measurement of the paper white has a diameter of CIExy = 0,001. This roughly correlates to DE00 = 1. The average of all 23 observers is not plotted, but is very close to the measurement of paper white.

It can be seen that most observers find a best visual match where the monitor exhibits less blue (CIEx) and less green (CIEy) then the paper white, when evaluated with a laboratory telespectroradiometer and the 2° standard observer. Although the average of all observers (not plotted) lies close to the measurement of the paper white. The observers were not tested for colour deficiency, but most observers have a long experience in the graphics art industry.

Trying to understand why some observers (WR53 = ID23, AD67=ID4, NU53=ID15 and CB78=ID7 and maybe JRKB70=ID3 and JH79=ID14) visual matches appear to be some sort of “outliers” compared to the majority it was assessed if these observers belong to either a group which show relative high variation and / or lower satisfaction with their visual “best match”. This would result in a higher uncertainty and could explain maybe partially the reported differences. If there is no correlation found it has to be assumed that the observer metamerism is source of the different visual matches.

Of those 7 observers 3 show less satisfaction with their visual match than the average and also show a relative high variation. One observer of the 7 shows a relative high variation and another observer is also less than average satisfied with the achieved visual match. So in sum 5 of the 7 “outliers” might be result of observers not seeing different than the other observers, but due lack of experience or time to achieve a repeatable good visual match.

To double check this theory it was evaluated which observers where very satisfied with their achieved visual match (“no colour difference”). This was true for observers ID5=APH86, ID8=DF55 and ID16=RB85. Observer with ID 5 and 8 showed little variation in chromaticity. Observer 8 is especially trained in doing visual comparisons. Both mentioned observers find a best visual match very close to the measurement of the paper white. Observer with ID16 was new in doing visual colour matches and might be less trained and easier to satisfy on the other side?

Conclusion

The results indicate that in this experiment experienced people could achive a good visual match, which also was close to the match the laboratory tele spectroradiometer would achive when using the 2° observer model. To achive reliable results in such an experiment, where small colour differences are of interest, preferably trained people should contribute to the experiments. Although some observers as well as the overall average of the observers came to a visual best match close to the agreement of the 2° standard observer colour metric the high number of observers chose a less bluish and greenish monitor match indicating that either individual observers which benefit from a colour metric different from the 2° standard observer or even that in general the “average” observer metamerism to colour measuring devices could be improved with a model different from the 2° observer. Further experiments need to be done with different monitors and spectral characteristics. Also using additional colours to match is needed to verify if systematic trends can be found.