ON THE COLOR RENDITION OF WHITE LIGHT SOURCES IN RELATION TO MEMORY PREFERENCE

DISSERTATION VON
SEBASTIAN MICHAEL BABILON

Vom Fachbereich Elektro- und Informationstechnik
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades
eines Doktors der Ingenieurwissenschaften
(Dr.-Ing.)

genehmigte Dissertation
von Sebastian Michael Babilon, M.Sc.
aus Lichtenfels

Erstgutachter: Prof. Dr.-Ing. habil. Tran Quoc Khanh
Zweitgutachter: Prof. Dr.-Ing. Edgar Dörsam

Tag der Einreichung: 26.06.2018
Tag der mündlichen Prüfung: 07.09.2018

Darmstadt 2018
D17

Sebastian Michael Babilon: *On the Color Rendition of White Light Sources in Relation to Memory Preference*

Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2018
Tag der mündlichen Prüfung: 07.09.2018

Dedicated to my beloved wife Laura.

Thank you for all your love and support
during the writing process
of this thesis.

ABSTRACT

Due to their potential use as an internal reference, memory colors have proven to provide an excellent conceptional approach for the color rendition evaluation of white light sources in terms of predicting visual appreciation. However, there are still some major drawbacks that can be identified in the principal design of existing memory-based or memory-related color quality metrics, of which the most severe is most likely that none of them were devised under realistic adaptation and viewing conditions. With the aim of contributing to a more comprehensive understanding of the nature of memory colors, a new experimental approach based on the color appearance rating of real familiar test objects perceived in a more realistic contextual viewing environment should therefore be presented as a main part of the current thesis trying to overcome the shortcomings of previous work. Besides attempting to draw universally valid conclusions about the memory colors' general characteristics, additional focus should be on the investigation of the impact of both the white point of adaptation and the observers' cultural background on the memory color assessments. By providing a comprehensive statistical analysis of the experimental data, it is shown that a significant effect on the observers' color appearance ratings can be reported for these two potential impact factors. With the corresponding dependencies being eventually known, a further goal of the current work should be the development of an improved memory-based color quality metric providing a superior tool for developers and manufacturers that can be used for the optimization of state-of-the-art lighting solutions in cases where visual appreciation and high user acceptability are more important than color fidelity. In order to validate the excellent predictive performance of this new metric proposal, the results of a comprehensive meta-correlation analysis based on the data of several different psychophysical studies are additionally reported. From this evaluation, it can be concluded that the newly proposed color quality metric outperforms all alternative approaches considered in the analysis making it an excellent choice to finally replace the CIE general color rendering index (R_a) in its use as an optimization criterion for modern light sources to achieve high visual appreciation and observer preference for which the latter has actually not been intended at all.

ZUSAMMENFASSUNG

Aufgrund ihrer per Definition gegebenen Eigenschaft als interne Bewertungsreferenz für die Beurteilung der Farberscheinung wohlvertrauter Objekte zu dienen, können Gedächtnisfarben prinzipiell für die farbmetrische Beurteilung der Farbwiedergabeeigenschaften weißer Lichtquellen in Bezug auf visuelles Gefallen herangezogen werden. Aktuell zur Verfügung stehende, gedächtnisfarbenbasierte Farbqualitätsmetriken weisen jedoch einige deutliche Defizite auf, wobei die Tatsache, dass keine dieser Metriken unter Berücksichtigung realistischer Adaptations- und Bebobachtungsbedingungen abgeleitet wurde, wahrscheinlich als das schwerwiegendste zu betrachten ist. Mit dem Ziel, ein umfassenderes Verständnis von der Natur der Gedächtnisfarben zu erlangen, wurde daher im Rahmen dieser Arbeit eine neue Probandenstudie zur Untersuchung der Bewertung der Farberscheinung realer, wohlvertrauter Objekte konzeptioniert und entsprechend durchgeführt. Zur Überwindung der angesprochenen Defizite bisheriger Arbeiten wurde den Probanden ein kontextbezogenes Stillleben in einem entsprechend möblierten, der Realität nachempfundenen Experimentierraum präsentiert, dessen Farberscheinung diese dann hinsichtlich ihrer Präferenz bewerten sollten. Abgesehen von dem Versuch, allgemeingültige Rückschlüsse über die grundsätzlichen Eigenschaften von Gedächtnisfarben abzuleiten, wurden zusätzlich sowohl der Einfluss der Adaptationsbedingungen als auch des kulturellen Hintergrunds der Probanden auf deren Bewertung der Gedächtnisfarben untersucht. Basierend auf der Durchführung einer umfassenden, statistischen Analyse konnte gezeigt werden, dass diese beiden Faktoren nachweislich einen signifikanten Einfluss auf die Farberscheinungsbewertung der Probanden haben. Unter Kenntnis der entsprechenden Abhängigkeiten wurde als weiteres Ziel dieser Doktorarbeit eine neue, auf Gedächtnisfarben basierende Farbqualitätsmetrik entwickelt, welche im Vergleich zu bestehenden Metriken, speziell in Fällen, in denen das visuelle Gefallen und die Nutzeraktzeptanz eine wichtigere Rolle spielen als das Erreichen einer hohen Farbtreue, sowohl für Entwickler als auch Hersteller ein deutlich besseres Werkzeug für die Optimierung von State-of-the-Art Lichtlösungen darstellt. Ausgehend von einer umfassenden Meta-Korrelationsanalyse, basierend auf den experimentellen Daten mehrerer psychophysikalischer Studien zur Thematik der Farbpräferenz, konnten die herausragenden Vorhersageeigenschaften der neuen Farbqualitätsmetrik nachgewiesen werden. Es konnte darüber hinaus gezeigt werden, dass dieser neue Vorschlag im direkten Vergleich besser abschneidet als alle möglichen Alternativen, die im Rahmen dieser Arbeit betrachtet wurden, so dass einer Ablösung der bisher (fälschlicherweise) weitesgehend verwendeten CIE R_a Metrik für die Optimierung von Lichtquellen zum Erreichen einer hohen Farbpräferenz durch die neuvorgeschlagene Farbqualitätsmetrik schlussendlich nichts mehr im Wege stehen sollte.

"To my friends, and family:
*You all may be batsh*t crazy,*
but even if I got to choose,
I'd still choose to be with you."

– Amelia Hutchins [1]

ACKNOWLEDGEMENTS

The work presented in this thesis was carried out during the years 2017–2018 at the Laboratory of Lighting Technology, Darmstadt as part of my postgraduate program intended to acquire a Doctoral degree in Electrical Engineering and Information Technology from the Technische Universität Darmstadt. With this acknowledgements section, I would like to explicitly thank those who made the emergence of this thesis possible by supporting me in the course of its creation process.

First of all, I owe my deepest gratitude to my supervisor Prof. Dr.-Ing. habil. Tran Quoc Khanh, with whom I was able to thoroughly discuss the topic and orientation of this research. His guidance into the field of color perception and rendition has been essential for the success of this work.

Furthermore, I am deeply grateful to Prof. Dr.-Ing. Edgar Dörsam for him declaring his agreement to appear as the second referee during the doctoral examination phase following the thesis' submission.

Special thanks goes to Dr. Vinh Trinh for his valuable support regarding the calculation of the various color quality metrics considered in this thesis. Moreover, I am indebted to many of my colleagues from the Laboratory of Lighting Technology, Darmstadt who always provided fruitful and very inspiring discussions on miscellaneous topics leading to a further broadening of my horizons beyond the scope of my own research.

At this point, I would also like to thank my physics mates Lars Jürgensen, Lewin Eidam, Julius Gronefeld, and Dirk Kulawiak for having shared the last eleven years of my life. May the force always be with you!

Last but not least, this thesis would not have been possible without the backing of my best friends Fabian Angermüller and Alexander Bruns, who both have been knowing me ever since I can remember, as well as without the loving support of my wife Laura, my parents Jutta and Michael Fischer and my sister Christina Rommel. You were always by my side for which I will be deeply grateful for the rest of my life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

BFL	Brown-Forsythe extension of Levene's test
BMS	Between-Targets Mean Square
CAM	Color Appearance Model
CAT	Chromatic Adaptation Transform
CAT02	A specific CAT adopted for the CIECAM02 definition
CCT	Correlated Color Temperature
CI	Confidence Interval
CIE	Commission Internationale de l'Éclairage
CIECAM02	Latest CAM officially published by the CIE
CIECAM02-UCS	Uniform three-dimensional color space based on the CIECAM02
CIEDE2000	Sophisticated color difference formula based on the CIELAB color space
CIELAB	Color space based on CIE ($L^*a^*b^*$) coordinates
CMF	Color Matching Function
CPI	Color Preference Index
CQS	Color Quality Scale
CRI	Color Rendering Index
CRI2012	Proposal for an updated CRI by Smet, Schanda, and Whitehead
CRT	Cathode Ray Tube
EMS	Residual Sum of Squares
ESD	Extreme Studentized Deviate
FCI	Feeling of Contrast Index
GAI	Gamut Area Index
HPE	Abbreviation for the Hunt-Pointer-Estévez cone response space
HS	Method of Hunter and Schmidt
ICC	Intraclass Correlation Coefficient
IES	Illuminating Engineering Society
IPT	Hue preserving color space developed by Ebner and Fairchild
IQR	Interquartile range
JMS	Between-Judges sum of Squares
JND	Just Noticeable Difference
KS	Kolmogorov-Smirnov test
LCD	Liquid Crystal Display
LED	Light Emitting Diodes
MCRI	Memory Color Rendition Index
PF	Performance Factor

PWM	Pulse Width Modulation
R, G, B	Color Matching Stimuli/Primaries
RGB	Red, green, and blue color channels
RGBA	Red, green, blue, and amber color channels
RGBWW	Red, green, blue, and warm white color channels
RMSE	Root-Mean-Square Error
SCD	Subjective Color Deviation
SPD	Spectral Power Distribution
SS	Sum-of-Squares
SSE	Sum of Squared Residuals
STRESS	Standardized Residual Sum of Squares
SW	Shapiro-Wilk test
TCS	Test Color Samples
TM	Technical Memoranda
UCS	Uniform Color Space

The following table is intended to give a comprehensive overview of the mathematical symbols used in the current thesis.

Symbol	Unit	Denotation
$\Delta\rho$	[–]	mean deviation of the spectral reflectance of an object from the theoretical reflectance of a perfectly white, diffusing standard
ΔC	[–]	memory-induced CIECAM02 chroma increment
$\Delta\overline{C}$	[–]	mean value of the memory-induced CIECAM02 chroma increments
$\Delta C'$	[–]	difference of the transformed CIELAB correlates of chroma
$\Delta E'$	[–]	CIECAM02-UCS color difference
$\Delta E'_{\text{chrom.}}$	[–]	CIECAM02-UCS chromatic color difference
$\Delta\overline{E'}_{\text{chrom.}}$	[–]	mean value of the CIECAM02-UCS chromatic color differences
ΔE_{00}	[–]	CIEDE2000 color difference
ΔE^*_{ab}	[–]	CIELAB color difference
$\Delta E_{f,i}$	[–]	difference between the chromaticities of the i^{th} test color sample illuminated by the test light source and its preferred chromaticities for a specific reference illumination
$\overline{\Delta E_{f,k}}$	[–]	arithmetic mean of the individually weighted differences $\Delta E_{f,i}$
ΔE_i	[–]	Euclidean distance in CIE 1964 color space between the i^{th} CRI test color sample rendered by the test and reference illuminant.
$\overline{\Delta E_k}$	[–]	arithmetic mean of the equally weighted differences $\Delta E_{k,i}$
$\Delta E_{k,i}$	[–]	difference between the chromaticities of Thornton's i^{th} test color sample illuminated by the test light source and its preferred chromaticities for a specific reference illumination
$\Delta H'$	[–]	modified difference of the transformed CIELAB correlates of hue
Δh	[°]	memory-induced CIECAM02 hue shift

$\Delta\overline{J'}$	[%]	deviation between the mean lightness value of the first and of one of the further experimental runs for a certain test object
$\Delta L'$	[–]	difference of the CIELAB correlates of lightness
ΔS_i	[–]	difference of the i^{th} test object's apparent and preferred chromaticities
Δuv	[–]	color difference in CIE 1960 (u, v) chromaticity diagram
$\Delta u'v'$	[–]	color difference in CIE 1976 (u', v') chromaticity diagram
$\Delta u_{f,i}$, $\Delta v_{f,i}$	[–]	preferred chromaticity shifts of the i^{th} test color sample as indicated by Judd
η^2	[–]	effect size of extra sum-of-squares F-test
$\eta_a(\lambda)$	[–]	original spectral reflectance of a certain test object during the first run of the experiments
$\eta_b(\lambda)$	[–]	spectral reflectance of a certain object candidate contemplated for one of the further runs of the experiments
θ	[–]	angular orientation of a chromatic tolerance ellipse defined in CIECAM02-UCS
λ	[nm]	wavelength of the emitted or perceived light
$\boldsymbol{\mu}_i$	[–]	vector of the CIECAM02-UCS coordinates of the centroid of the fitted Gaussian function assigned to the i^{th} familiar test object
ρ, γ, β	[–]	color coordinates of the color stimulus in the Hunt-Pointer-Estévez cone response space
ρ_{w}, γ_{w}, β_{w}	[–]	color coordinates of the device or scene white in the Hunt-Pointer-Estévez cone response space
ρ_{a}, γ_{a}, β_{a}	[–]	color coordinates of the color stimulus in the Hunt-Pointer-Estévez cone response space after applying non-linear compression and luminance-level adaptation
$\rho_{\text{w,a}}$, $\gamma_{\text{w,a}}$, $\beta_{\text{w,a}}$	[–]	color coordinates of the device or scene white in the Hunt-Pointer-Estévez cone response space after applying non-linear compression and luminance-level adaptation
ϱ	[–]	Pearson correlation coefficient
$\overline{\varrho}$	[–]	average Pearson correlation coefficient
$\boldsymbol{\Sigma}_i$	[–]	covariance matrix of the i^{th} familiar test object
σ	[–]	standard deviation
$\hat{\tau}^2$	[–]	heterogeneity estimator

$\hat{\tau}^2_{\mathrm{HS}}$	[–]	heterogeneity estimator as defined by Hunter and Schmidt
$\phi(\lambda)$	[a.u.]	spectral color stimulus
A	[–]	area size of a chromatic tolerance ellipse defined in CIECAM02-UCS
A^*	[–]	achromatic response of the color stimulus
A^*_{w}	[–]	achromatic response of the device or scene white
a, b	[–]	semi-major and semi-minor axis of a chromatic tolerance ellipse defined in CIECAM02-UCS
a^*	[–]	CIELAB component representing the green-red perception
a_{02}	[–]	CIECAM02 color-difference signal for the green-red perception
$a_{1,i}, ..., a_{7,i}$	[–]	fitting parameters defining the size, shape, orientation, and location of the bivariate Gaussian function assigned to the i^{th} familiar test object
a_{CPI}, b_{CPI}	[–]	scaling constants of the CPI
a_i	[–]	attenuation factor representing statistical bias of the i^{th} lighting scenario
b^*	[–]	CIELAB component representing the blue-yellow perception
b_{02}	[–]	CIECAM02 color-difference signal for the blue-yellow perception
b_{CRI}	[–]	scaling constant of the CRI
C	[–]	CIECAM02 perceptual correlate of chroma
C^*_{ab}	[–]	CIELAB correlate of chroma
C'_i	[–]	transformed CIELAB correlates of chroma for the CIEDE2000 calculation
c, N_c, F	[–]	model parameters of the categorical surround conditions in CIECAM02
c_1	[W m^2]	first radiation constant
c_2	[K m]	second radiation constant
D	[–]	parameter to set the degree of adaptation in CIECAM02
D^2	[–]	effect size for the multivariate case
d	[–]	effect size as defined by Cohen for the univariate case
df	[–]	degrees of freedom
E	[lx]	illuminance
$\overline{E}$	[lx]	mean illuminance

e_t	[–]	CIECAM02 eccentricity factor
$f(T)$	[–]	objective function to be minimized for the CCT calculation
$f_i(\mathbf{x}_i)$	[–]	bivariate Gaussian function to be fitted to the observers' color appearance rating data for the i^{th} familiar test object
$g_{kl,i}$	[–]	ellipse parameters defining the size, shape, and orientation of Sanders' i^{th} tolerance ellipse
H	[–]	CIECAM02 hue quadrature
H_0	[–]	null hypothesis
h	[°]	CIECAM02 hue angle
h_{ab}	[°]	CIELAB correlate of hue
h'_i	[°]	transformed CIELAB correlates of hue for the CIEDE2000 calculation
$\text{ICC}(2,k)$	[–]	two-way random effects intraclass correlation coefficient as defined by Shrout and Fleiss
J	[–]	CIECAM02 perceptual correlate of lightness
J', a'_M, b'_M	[–]	color coordinates of the CIECAM02-UCS
$\overline{J'}$	[–]	mean lightness correlate of the CIECAM02-UCS
K_m	[$\text{lm}\,\text{W}^{-1}$]	luminous efficacy of radiation
k_Chinese, k_German	[–]	sample size of the Chinese and German variability distributions
k_L, k_C, k_H	[–]	parametric correction factors to be used in the CIEDE2000 formula
L	[$\text{cd}\,\text{m}^{-2}$]	luminance
L, U	[–]	lower and upper 95 % CI bounds for the difference $\hat{\bar{r}}_{\text{c},i} - \hat{\bar{r}}_{\text{c},j}$ of two different color quality metrics i and j
$\overline{L}$	[$\text{cd}\,\text{m}^{-2}$]	mean luminance
L^*	[–]	CIELAB correlate of lightness
L_A	[$\text{cd}\,\text{m}^{-2}$]	luminance of the adapting field
L_b	[$\text{cd}\,\text{m}^{-2}$]	background luminance
L_DW	[$\text{cd}\,\text{m}^{-2}$]	luminance of the device or scene white
L_SW	[$\text{cd}\,\text{m}^{-2}$]	luminance of the surround white
l_i, u_i	[–]	lower and upper 95 % CI bounds of the artifact-corrected $\hat{\bar{r}}_{\text{c},i}$ value obtained for the i^{th} color quality metric
$\bar{l}(\lambda)$, $\bar{m}(\lambda)$, $\bar{s}(\lambda)$	[–]	cone fundamentals of L-, M-, and S-cones
M	[–]	CIECAM02 perceptual correlate of colorfulness
$M(\lambda, T)$	[$\text{W}\,\text{m}^{-2}\,\text{nm}^{-1}$]	spectral radiant exitance of a black body at given temperature T in Kelvin

M'	[–]	CIECAM02-UCS colorfulness
M_{CAT02}	[–]	linear CAT02 chromatic adaptation transform matrix
M_{HPE}	[–]	linear transformation matrix for conversion to Hunt-Pointer-Estévez cone response space
M_1, M_2	[–]	chromaticity-dependent scaling factors for calculating the CIE daylight phases
$\overline{N}$	[–]	average observer number of all studies used for the meta-correlation analysis
N_{bb}, N_{cb}, F_{L}	[–]	CIECAM02 background and adapting field parameters
N_i	[–]	number of observers rating the i^{th} lighting scenario
$N_i^{\text{opt.}}$	[–]	optimal study weights
n_{Chinese}, n_{German}	[–]	sample sizes of the bivariate similarity distributions of the Chinese and German observer group
P	[–]	correlate of the red-green perception of the IPT color space
PF/3	[–]	inter-(intra-)observer performance factor defined by Guan and Luo for describing the inter-(intra-)observer variability in color appearance rating or color matching experiments
p_1^*, p_2^*, p_3^*	[–]	scaling parameters used for defining the MCRI
p_i	[–]	individual weighting factor of the i^{th} familiar test object used for the R_{MCPI} definition
Q	[–]	CIECAM02 perceptual correlate of brightness
Q_{a}, Q_{f}, Q_{p}, Q_{g}	[–]	various measures of the CQS metric
R, G, B	[–]	sharpened spectral sensitivity responses calculated from the tristimulus values of a color stimulus by applying the CAT02 transform
R^2	[–]	adjusted coefficient of determination
$R(\lambda)$	[–]	spectral reflectance
R_a	[–]	general CRI
R_{CPI}	[–]	Thornton's color preference index
R_{c}, G_{c}, B_{c}	[–]	adapted spectral sensitivity responses of the color stimulus in CIECAM02
R_{f}	[–]	general color fidelity index
R_{g}	[–]	gamut-based measure of the IES TM-30 metric
$R_{\text{flatt.}}$	[–]	Judd's flattery index
R_i	[–]	special CRI calculated for the i^{th} test color sample

R_{MCPI}	[–]	newly proposed memory color preference index
R_{m}	[–]	Smet's memory color rendition index
R_{p}	[–]	Sanders' preferred color index
R_{p}, G_{p}, B_{p}	[–]	amounts of the red, green, and blue primaries required to match an arbitrary test stimulus
R_T	[–]	rotation parameter to correct for distortions of the CIELAB color space in the blue region during the CIEDE2000 calculation
R_{w}, G_{w}, B_{w}	[–]	sharpened spectral sensitivity responses calculated from the tristimulus values of the device or scene white by applying the CAT02 transform
$R_{\mathrm{w,c}}$, $G_{\mathrm{w,c}}$, $B_{\mathrm{w,c}}$	[–]	adapted spectral sensitivity responses of the device or scene white in CIECAM02
RMSE	[–]	root-mean-square error
$\mathrm{RMSE}_{\mathrm{adj.}}$	[–]	adjusted root-mean-square error
$\overline{r}$	[–]	sampling error corrected, weighted average Spearman correlation coefficient
$\hat{\overline{r}}$	[–]	unbiased estimator of the true correlation
$\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$	[–]	color matching functions as obtained from the results of Wright and Guild
$r_{\theta,i}$	[–]	radius of the tolerance ellipse of Sanders' i^{th} test object measured in the same direction as the vectorial orientation of ΔS_i
$r_{0,i}$	[–]	individual Spearman correlation coefficient of the i^{th} lighting scenario after correction for restrictions in range
$\hat{\overline{r}}_{\mathrm{c}}$	[–]	best estimate of the true correlation between a metric's prediction and the observers' preference ratings obtained from a completely artifact-corrected meta-correlation analysis
r_i	[–]	individual Spearman correlation coefficient of the i^{th} lighting scenario
$r_i^{\mathrm{corr.}}$	[–]	individual Spearman correlation coefficient of the i^{th} lighting scenario after correction for inter-observer idiosyncrasy
r_i^{bias}	[–]	individual Spearman correlation coefficient of the i^{th} lighting scenario after correction for statistical bias
$\sqrt{r_{yy,i}}$	[–]	degree of attenuation due to idiosyncrasy in the observers' preference ratings for the i^{th} lighting scenario
$S(\lambda)$	[a.u.]	(relative) spectral power distribution

$S_0(\lambda)$, $S_1(\lambda)$, $S_2(\lambda)$	[–]	standardized spectral components for calculating the relative SPDs of the CIE daylight phases
S_a	[–]	general degree of similarity
$S_i(P_i, T_i)$	[–]	similarity distribution function of the i^th familiar test object adopted by Smet *et al.*
$S_i(\mathbf{x}_i)$	[–]	similarity distribution function of the i^th familiar test object
S_L, S_C, S_H	[–]	lightness, chroma, and hue weighting functions for the calculation of the CIEDE2000 color difference
S_R	[–]	surround ratio
SCD_a	[–]	general subjective color deviation calculated by averaging the individual SCD_i values
SCD_i	[–]	subjective color deviation given by the ratio of ΔS_i and $r_{\theta,i}$
SSE	[–]	sum of squared residuals
STRESS	[–]	standardized residual sum-of-squares
s	[–]	CIECAM02 perceptual correlate of saturation
T	[K]	temperature of the Planckian radiator
T	[–]	correlate of the yellow-blue perception of IPT color space
$T(\lambda)$	[–]	spectral transmittance
T_cp	[K]	CCT of the test illuminant
u, v	[–]	coordinates of the CIE 1960 uniform chromaticity scale
u', v'	[–]	coordinates of the CIE 1976 uniform chromaticity scale
u_k, v_k	[–]	CIE 1960 (u, v) coordinates of an arbitrary test illuminant
$u_{\mathrm{k},i}$, $v_{\mathrm{k},i}$	[–]	CIE 1960 (u, v) coordinates of the i^th test color sample as illuminated by the test light source
$u'_{\mathrm{k},i}$, $v'_{\mathrm{k},i}$	[–]	von-Kries-transformed CIE 1960 (u, v) coordinates of the i^th test color sample accounting for the two different adaptation conditions under test and reference light source during CRI calculation
$u_\mathrm{P}(T)$, $v_\mathrm{P}(T)$	[–]	CIE 1960 (u, v) coordinates of the Planckian radiator at temperature T
u_r, v_r	[–]	CIE 1960 (u, v) coordinates of the reference illuminant
$u_{\mathrm{r},i}$, $v_{\mathrm{r},i}$	[–]	CIE 1960 (u, v) coordinates of the i^th test color sample as illuminated by the reference light source

U_x	[–]	inverse of the range restriction parameter
u_x	[–]	range restriction parameter
$V(\lambda)$	[–]	photopic spectral luminous efficiency function
$w(\lambda)$	[–]	simple filter function for the calculation of $\mathrm{RMSE}_{\mathrm{adj.}}$
W^*, U^*, V^*	[–]	color coordinates of the CIE 1964 color space
X, Y, Z	[–]	tristimulus values of the 2° standard observer
X_{10}, Y_{10}, Z_{10}	[–]	tristimulus values of the 10° standard observer
X_n, Y_n, Z_n	[–]	tristimulus values of a reference white for calculating CIELAB coordinates
X_w, Y_w, Z_w	[–]	tistimulus values of the device or scene white in CIECAM02
x, y, z	[–]	CIE 1931 chromaticity coordinates
$\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$	[–]	color matching functions of the CIE 1931 2° standard colorimetric observer
x_{10}, y_{10}, z_{10}	[–]	CIE 1964 chromaticity coordinates
$\overline{x}_{10}(\lambda)$, $\overline{y}_{10}(\lambda)$, $\overline{z}_{10}(\lambda)$	[–]	color matching functions of the CIE 1964 10° standard colorimetric observer
$x_\mathrm{D}(T_\mathrm{cp})$, $y_\mathrm{D}(T_\mathrm{cp})$	[–]	CIE 1931 chromaticity coordinates of the daylight locus at a given T_cp
$\mathbf{x}_i$	[–]	vector of CIECAM02-UCS chromaticities of the i^{th} familiar test object
$x_{\mathrm{p},i}$, $y_{\mathrm{p},i}$	[–]	preferred CIE 1931 chromaticities of the i^{th} test object used in the experiments of Sanders
$x_{\mathrm{t},i}$, $y_{\mathrm{t},i}$	[–]	apparent CIE 1931 chromaticities of the i^{th} test object used in the experiments of Sanders
Y_b	[–]	background luminous factor

Part 1

INTRODUCTION

"The whole of science is nothing more than a continuous refinement of everyday thinking"

– Albert Einstein [2]

1 INTRODUCTION

Since the advent of the first commercially available incandescent light bulbs more than 130 years ago [3, 4], fuel-based light sources such as candles, oil lamps, and gas lanterns got more and more replaced by electrically driven illuminants. Soon, these new light sources became an indispensable part of people's daily lives and a driving force in the evolution of our modern society, even though lighting-related research in the early days was only about increasing luminous efficiency and life time [5–7]. This, however, drastically changed with the commercialization of fluorescent gas-discharge lamps by the end of the 1930s. Being based on the physical principle of converting ultraviolet radiation into visible light via the application of fluorescent coating materials, these light sources offer – in contrast to incandescent light bulbs – greatly enhanced spectral flexibility. Whereas the emitted light spectra of incandescent light sources are more or less fixed with correlated color temperatures of approximately 2700 K, the spectra of fluorescent lamps can be varied over a wide range of possibilities simply by altering the mixture of the fluorescent coatings inside the lamp tube [8, 9]. Hence, fluorescent gas-discharge lamps can be tailored to almost any desired color temperature ranging from warm-white at 2700 K through neutral-white at 4000 K to cool-white at 4500 K. To some extent, even the daylight phases at 5000 or 6500 K, resulting in a more bluish-white, can be spectrally imitated with these kind of light sources [10–12].

The new spectral flexibility in designing light emitting devices entailed an increased demand for advanced research on spectral optimization for improving the color appearance of illuminated objects. At the same time, first attempts were made to evaluate and quantify the ability of artificial light sources to render object colors [13–15]. Based on these efforts, the CIE published in 1965 a recommendation to evaluate the color rendering properties of light sources by calculating an average color rendering index (CRI) R_a using a test sample method [16, 17], which mathematically compares the color appearance of some test samples illuminated by a test source to the color appearance of the same samples illuminated by a reference source. By definition, a value of $R_a = 100$ is attributed to the reference light source and the maximum value a test light source can achieve in case that the calculated color differences of the test samples under reference and test illuminant are negligibly small.

However, as the experienced reader may know, a high CRI value as an objective measure does not necessarily mean high end-user acceptance and visual appreciation [18–22]. In general lighting applications (e.g., home, office, factory, restaurant,...), in the shop and retail sector as well as in the film industry one is usually more interested in the perceived color quality of the lighting [23], which is a quite subjective aspect of how appealing real objects look like for the observer. In this context, several former studies revealed that preferred object colors tend to require an additional amount of saturation compared to the actual object colors under reference illumination [24–31]. In order to evaluate the degree of preference to which an artificial illuminant renders object colors, some additional metrics like Sanders' preferred color index [26], Judd's flattery index [27], and Thornton's color preference index [30] have been defined, all of which, in contrast to the CRI, focus on the more subjective aspects of lighting color quality.

Besides color preference, another concept has been introduced to assess the color quality of artificial light sources – the concept of memory colors. In psychology, the term "memory color" has been known since the late 19th century. Following the definition of Hering [32],

it describes *"the color in which we have most consistently seen the external object"* and which, therefore, is *"indelibly impressed on our memory"*. In other words, the term "memory color" is used to describe the color appearance an individual observer has in mind when thinking of or looking at certain familiar objects, i.e., objects that are frequently visually experienced in the course of a lifetime. Even though this creates a relatively stable but also very personal standard on how the color of familiar objects should look like for the individual observer, memory colors of different observers are typically more general in their characteristics as one might expect at first glance [33].

First experiments aiming to quantify the human assessment of memory colors were conducted by Newhall and Pugh [34] published in 1957 and by Bartleson [35] published in 1960. Both experiments investigated the memory colors of familiar objects (e.g., green grass, skin, red brick, blue sky,...) using a large number of Munsell color patches which were presented to the observer under controlled experimental conditions. The results of both experiments indicated that memory colors often tend to be exaggerated towards the typical or dominant hues commonly associated with these objects, i.e., the color of grass is recalled greener, the color of a brick is recalled redder, etc. than their average measured colors [24, 36]. Furthermore, the data showed a tendency of the memory colors – similiar to those of the preferred colors – to appear more saturated than the actual object colors, an overall trend which has been confirmed in various subsequent experiments [31, 33, 37–41].

In recent years, both memory and preferred colors have extensively been studied in various fields of color research, such as color constancy [42–44], color reproduction [33, 39, 45–51], and of course color rendering [23, 26, 27, 30, 52–54]. As stated by Smet *et al.* [55], memory colors basically provide cues to the visual system which may serve as an internal reference in the assessment of color appearance and color quality. In addition, they have also been suggested to notably contribute to the mechanisms of color constancy by guiding the visual system in the process of assigning suitable colors to objects irrespective (at least to some degree) of the perceived illuminant [43, 44].

Hence, indication is given that memory and preferred colors may provide very powerful concepts in the evaluation of color rendering properties of white light sources beyond the scope of color fidelity which emphasize the more subjective aspects of color quality that are responsible for the preference of a perceived illuminated scene. For this reason, the current thesis should mainly focus on the application of both concepts in the context of color rendition evaluation in terms of attractiveness and visual appreciation. Based on the work of Smet *et al.* [23, 40, 53, 54], it can be concluded that with their ability to serve as an internal reference in the assessment of color appearance and color quality, memory colors, in particular, play a crucial role in the rating of visual appreciation of modern light sources, which conceptionally resulted in the creation of the first and so far only purely memory-based color quality metric intended to replace the CRI in applications where achieving a high fidelity is not the only goal. However, their research on closer examination also revealed some major drawbacks and shortcomings which will be discussed in a later chapter of this thesis. The intention of this work, therefore, is to develop a new memory-based color quality metric for the evaluation of the color rendition of arbitrary white light sources which addresses the shortcomings of previous definitions and exhibits a significantly better performance in predicting observers' preference ratings.

In order to achieve this goal while keeping the reading of this thesis as comprehensible as possible, the individual chapters are organized as follows. In Chapter 2, an introduction to the basic CIE colorimetry as applied in the subsequent chapters will be given. The presented theory should help the reader to get a better and faster access to the achieved research

results. Based on these considerations, various methods of specifying the color rendering properties of white light sources will be presented in Chapter 3, where the focus of discussion should be put on the existing preference- and memory-based color quality metrics of Sanders, Judd, Thornton, and Smet *et al.*. Chapter 4 is one of two main parts of the current thesis. This chapter summarizes the results of several psychophysical experiments conducted by the author to investigate how human beings assess memory colors of real familiar objects under realistic immersive viewing and adaptation conditions. Furthermore, it should be figured out whether or not variations in the adaptation conditions and the cultural background of the observers have a significant impact on the memory color assessments. Based on the reported findings, an updated and improved memory-based color quality metric should be presented in Chapter 5, which is the second main part of the current thesis. In order to prove that this newly proposed metric outperforms all alternative approaches, a meta-correlation analysis including experimental data of several different visual studies investigating color preference ratings of human observers in both metameric and multi-CCT lighting scenarios will be performed. Finally, the thesis will be completed in Chapter 6 with some concluding remarks and a comprehensive outlook on future research intentions.

REFERENCES OF THE INTRODUCTION

[3] J. Levy, *Really Useful: The Origins of Everyday Things*. Firefly Books, Berlin, Germany, 2002.

[4] T. A. Edison, "Electric-Lamp," Patent US 223898 A, Jan. 1880.

[5] L. H. Latimer, "Process of Manufactoring Carbons," Patent US 252386 A, Jan. 1882.

[6] C. B. Davenport, "The Nernst Lamp," *Science*, vol. 8, no. 203, p. 689, 1898.

[7] L. R. Ingersoli, "On the Radiant Efficiency of the Nernst Lamp," *Physical Review*, vol. 17, no. 5, p. 371, 1903.

[8] T. Welker, "Recent Developments on Phosphors for Fluorescent Lamps and Cathode-Ray Tubes," *Journal of Luminescence*, vol. 48–49, no. 1, p. 49, 1991.

[9] A. M. Srivastava and T. Sommerer, "Fluorescent Lamp Phosphors," *Interface*, vol. 7, no. 2, p. 28, 1998.

[10] F. Grum, "Artificial Light Sources for Simulating Natural Daylight and Skylight," *Applied Optics*, vol. 7, no. 1, p. 183, 1968.

[11] S. J. Dain, "Daylight Simulators and Colour Vision Tests," *Ophthalmic and Physiological Optics*, vol. 18, no. 6, p. 540, 1998.

[12] H. Xu, M. R. Luo, and B. Rigg, "Evaluation of Daylight Simulators Part I: Colorimetric and Spectral Variations," *Coloration Technology*, vol. 119, no. 2, p. 59, 2003.

[13] D. Nickerson, "Measurement and Specification of Color Rendition Properties of Light Sources," *Illuminating Engineering*, vol. 53, no. 2, p. 77, 1958.

[14] J. L. Ouweltjes, "The Specification of Colour Rendering Properties of Fluorescent Lamps," *Die Farbe*, vol. 9, p. 207, 1960.

[15] D. Nickerson, "Light Sources and Colour rendering," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 57, 1960.

[16] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.1, 1965.

[17] D. Nickerson and C. W. Jerome, "Color Rendering of Light Sources: CIE Method of Specification and its Application," *Illuminating Engineering*, vol. 60, no. 4, p. 262, 1965.

[18] S. Jost-Boissard and M. Fontoynont, "Optimization of LED-based Light Blendings for Object Presentation," *Color Research & Application*, vol. 34, no. 4, p. 310, 2009.

[19] W. Davis and Y. Ohno, "Color quality scale," *Optical Engineering*, vol. 49, no. 3, p. 033 602, 2010.

[20] M. S. Islam, R. Dangol, M. Hyvärinen, *et al.*, "Investigation of User Preferences for LED Lighting in Terms of Light Spectrum," *Lighting Research & Technology*, vol. 45, no. 6, pp. 641–665, 2013.

[21] R. Dangol, M. Islam, M. Hyvärinen, *et al.*, "Subjective Preferences and Colour Quality Metrics of LED Light Sources," *Lighting Research & Technology*, vol. 45, p. 666, 2013.

[22] R. R. Baniya, R. Dangol, P. Bhusal, *et al.*, "User-acceptance Studies for Simplified Light-Emitting Diode Spectra," *Lighting Research & Technology*, vol. 47, p. 177, 2015.

[23] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Memory Colours and Colour Quality Evaluation of Conventional and Solid-State Lamps," *Optics Express*, vol. 18, no. 25, pp. 26 229–26 244, 2010.

[24] G. B. Buck and H. C. Froelich, "Color Characteristics of Human Complexions," *Illuminating Engineering*, vol. 43, no. 1, p. 27, 1948.

[25] C. L. Sanders, "Color Preferences for Natural Objects," *Illuminating Engineering*, vol. 54, p. 452, 1959.

[26] C. L. Sanders, "Assessment of Color Rendition under an Illuminant Using Color Tolerances for Natural Objects," *Illuminating Engineering*, vol. 54, p. 640, 1959.

[27] D. B. Judd, "A Flattery Index for Artificial Illuminants," *Illuminating Engineering*, vol. 62, p. 593, 1967.

[28] C. W. Jerome, "Flattery vs. Color Rendition," *Journal of the Illuminating Engineering Society*, vol. 1, no. 3, p. 208, 1972.

[29] C. W. Jerome, "The Flattery Index," *Journal of the Illuminating Engineering Society*, vol. 2, no. 4, p. 351, 1973.

[30] W. A. Thornton, "A Validation of the Color-Preference Index," *Journal of the Illuminating Engineering Society*, vol. 4, no. 1, p. 48, 1974.

[31] P. Siple and R. M. Springer, "Memory and Preference for the Colors of Objects," *Perception & Psychophysics*, vol. 34, no. 4, p. 363, 1983.

[32] E. Hering, *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer Verlag, 1920.

[33] C. J. Bartleson, "Color in Memory in Relation to Photographic Reproduction," *Photographic Science and Engineering*, vol. 5, no. 6, p. 327, 1961.

[34] S. M. Newhall, R. W. Burnham, and J. R. Clark, "Comparison of Successive with Simultaneous Color Matching," *Journal of the Optical Society of America*, vol. 47, no. 1, p. 43, 1957.

[35] C. J. Bartleson, "Memory Colors of Familiar Objects," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 73, 1960.

[36] C. D. Hendley and S. Hecht, "The Colors of Natural Objects and Terrains, and their Relation to Visual Color Deficiency," *Journal of the Optical Society of America*, vol. 39, no. 10, p. 870, 1949.

[37] J. Pérez-Carpinell, M. D. de Fez, R. Baldoví, *et al.*, "Familiar Objects and Memory Color," *Color Research & Application*, vol. 23, no. 6, pp. 416–427, 1998.

[38] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Representation of Memory Prototype for an Object Color," *Color Research & Application*, vol. 24, no. 6, pp. 393–410, 1999.

[39] P. Bodrogi and T. Tarczali, "Colour Memory for Various Sky, Skin, and Plant Colours: Effect of the Image context," *Color Research & Application*, vol. 26, no. 4, p. 278, 2001.

[40] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Colour Appearance Rating of Familiar Real Objects," *Color Research & Application*, vol. 36, no. 3, pp. 192–200, 2011.

[41] T. Tarczali, S. Du Park, P. Bodrogi, *et al.*, "Long-term Memory Colors of Korean and Hungarian Observers," *Color Research & Application*, vol. 31, no. 3, p. 176, 2006.

[42] A. C. Hurlbert and Y. Ling, "If It's a Banana, It must be Yellow: The Role of Memory Colors in Color Constancy," *Journal of Vision*, vol. 5, no. 8, p. 787, 2005.

[43] T. Hansen, M. Olkkonen, S. Walter, *et al.*, "Memory Modulates Color Appearance," *Nature Neuroscience*, vol. 9, no. 11, p. 1367, 2006.

[44] J. J. M. Granzier and K. R. Gegenfurtner, "Effects of Memory Color on Color Constancy for Unknown Colored Objects," *i-Perception*, vol. 3, no. 3, p. 190, 2012.

[45] P. Bodrogi and T. Tarczali, "Investigation of Colour Memory," in *Colour Image Science: Exploiting Digital Media*, L. W. MacDonald and M. R. Luo, Eds. Chichester: John Wiley & Sons Limited, 2002, pp. 23–48.

[46] H. Zeng and R. Luo, "Modelling Memory Color Region for Preference Color Reproduction," in *Proceedings of the SPIE 7528, Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*, San Jose, CA, USA: International Society for Optics and Photonics, 2010, p. 752 808.

[47] S. Xue, M. Tan, A. McNamara, *et al.*, "Exploring the Use of Memory Colors for Image Enhancement," in *Proceedings of the SPIE 9014, Human Vision and Electronic Imaging XIX*, San Francisco, CA, USA: International Society for Optics and Photonics, 2014, p. 901 411.

[48] C. Boust, H. Chahine, M. B. Chouikha, *et al.*, "Color Correction Judgements of Digital Images by Experts and Naive Observer," in *Proceedings of the PICS Conference 2003*, Rochester, NY, USA: Society for Imaging Science and Technology (IS&T), 2003, pp. 4–9.

[49] C. Boust, H. Brettel, F. Viénot, *et al.*, "Color Enhancement of Digital Images by Experts and Preference Judgements by Observers," *Journal of Imaging Science and Technology*, vol. 50, no. 1, pp. 1–11, 2006.

[50] C. Boust, F. Cittadini, M. B. Chouikha, *et al.*, "Does an Expert Use Memory Colors to Adjust Images?" In *Proceedings of the 12th Color and Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 2004, pp. 347–353.

[51] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Color Reproduction and the Naturalness Constraint," *Color Research & Application*, vol. 24, no. 1, pp. 52–67, 1999.

[52] K. Smet, S. Jost-Boissard, W. R. Ryckaert, *et al.*, "Validation of a Colour Rendering Index Based on Memory Colours," in *Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2010, pp. 136–142.

[53] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "A Memory Colour Quality Metric for White Light Sources," *Energy and Buildings*, vol. 49, p. 216, 2012.

[54] K. Smet and P. Hanselaer, "Memory and Preferred Colours and the Colour Rendition of White Light Sources," *Lighting Research & Technology*, vol. 48, no. 4, p. 393, 2016.

[55] K. A. G. Smet, Y. Lin, B. V. Nagy, *et al.*, "Cross-Cultural Variation of Memory Colors of Familiar Objects," *Optics Express*, vol. 22, no. 26, p. 32 308, 2014.

[56] H. Matisse, *Quotes of Henri Matisse*, `https://quotefancy.com/quote/1315121/Henri-Matisse-Color-was-not-given-to-us-in-order-that-we-imitate-Nature-It-was-given-to`, [Online; accessed 22-May-2018].

Part 2

INTRODUCTION TO BASIC COLORIMETRY

"Color was not given to us in order that we imitate nature. It was given to us so that we can express our emotions"

– Henri Matisse [56]

2

INTRODUCTION TO BASIC COLORIMETRY

This chapter gives an introduction to the basics of CIE colorimetry which should help the reader to get a better and faster conceptual access to the research results presented in the subsequent parts of this thesis. Special emphasis should be put on the psychophysical concept of color from a lighting engineering point of view. In this context, the aim of the CIE colorimetry is to derive a universally valid mathematical description of the way how human beings perceive colors of illuminated objects in a complex and realistic scenery. Due to significant inter-personal variations and the general complexity of the cognitive signal processing in the human brain involved in the perception of color, this of course is a very challenging task which has not yet been solved in an entirely satisfying manner. Nevertheless, significant progress has been made since the CIE's first attempts to cast the human perception of colors into formulas resulting in a quite huge variety of sophisticated color appearance models and perceptual uniform color spaces. Starting with the fundamental CIE colorimetry of the 1930s, the following sections should give an overview of this progress from the early beginnings to the latest color appearance model recommended by the CIE.

2.1 FUNDAMENTAL CIE COLORIMETRY

The concept of color can be conceived as a sensation caused by neural stimuli of the photoreceptor cells of the human eye which are processed in specific regions of the human brain eventually leading to something one would simply call a color impression. The reason for this neural stimulation is the incidence of light through the pupil of the eye and onto the retina which basically contains three different types of photoreceptor cells being responsible for color vision [57–61]. These are the so-termed L-, M-, and S-cones which stands for long, middle, and short wavelength, respectively, designated to characterize the maximum of their spectral sensitivities. These also called cone fundamentals $\bar{l}(\lambda)$, $\overline{m}(\lambda)$, and $\bar{s}(\lambda)$ which are illustrated in Fig. 2.1 for descriptive purposes can either be determined from the outcome of certain color matching experiments [62–64] or by direct microspectrophotometric measurements on the human eye [65–67]. As stated by the CIE [64], the cone fundamentals basically depend on both the effective field of view and the age of the individual observer so that corresponding formulas for their calculation were given and should be used where applicable.

When light is absorbed in the cones of the retina, their photosensitive pigments are excited leading to variations in the electrical potential which are processed by a series of relay cells and eventually transmitted through a nerve fiber to the human brain. On their way, the signals of the different cone types are interconnected by neurons and encoded for transmission resulting in one achromatic $(L + M)$ and two chromatic signals $(L - M)$ and $((L + M) - S)$ that have to be interpreted in the brain cortex [57, 68]. The two chromatic signals basically define the red-green and blue-yellow perception, respectively. In conjunction with their achromatic counterpart, the neural processing of these signals finally leads to the perceptual color attributes of brightness, hue, and colorfulness which determine the perceived color impression. According to Hunt [68], these attributes can be comprehend as follows:

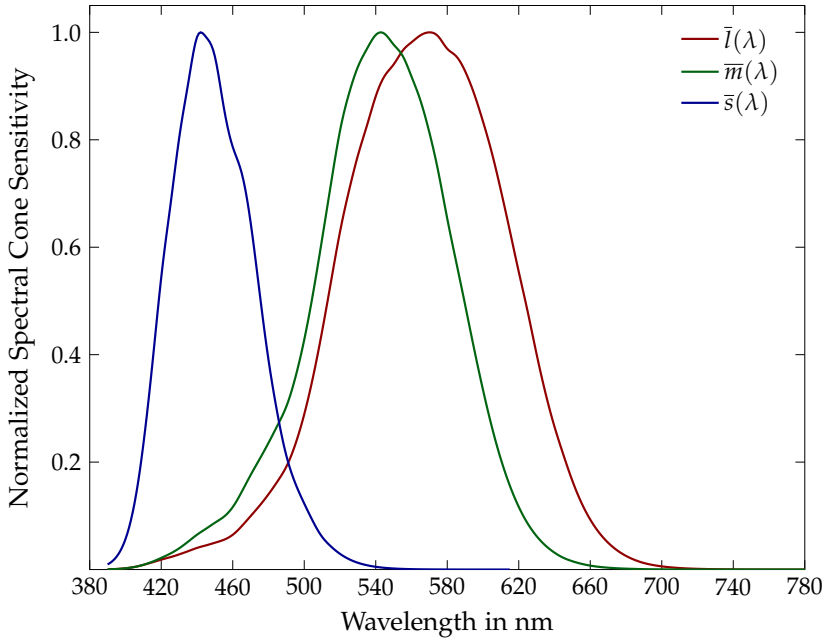


Figure 2.1 – Normalized cone sensitivities of the human eye measured as a function of wavelength which create the basis of photopic vision. Please note that the cone fundamentals basically depend on both the effective field of viewing and the age of the individual observer. Hence, the curves plotted here are for descriptive purposes only and were reproduced from the data provided by Stockman *et al.* [62, 63].

BRIGHTNESS Attribute of a visual perception according to which an area appears to exhibit more or less light so that the adjectives bright or dim can be associated with it.

HUE Attribute of a visual perception according to which an area appears to be similar to one, or to proportions of two, of the perceived colors red, yellow, green, and blue.

COLORFULNESS Attribute of a visual perception according to which an area appears to exhibit more or less of its hue.

Hence, the mathematical evaluation of color vision from known spectral power distributions (SPDs) of light sources resulting in a quantification of the perceptual attributes can in principle be described on basis of the cone fundamentals. However, from a historical point of view it was impossible to measure the spectral sensitivities of the cones and their distribution on the retina with the necessary precision so that instead trichromatic color matching was used to create a proper basis for a system of color measurement in order to mathematically describe the perception of color [69].

As can be seen from Fig. 2.2, which schematically shows the corresponding experimental setup, a monochromatic reference stimulus produced by illuminant C and a compound color as a result of the mixture of three monochromatic matching stimuli R, G, and B are seen on each half of a specific test field presented to the observer. In general, both halves exhibit a differently colored but – due to the diffusers 1 and 2 – uniform appearance when viewed through the optical system of the color matching apparatus which, of course, is much more complex in reality than the one sketched here for illustrative purposes only. In any case, the task is to match the color impression of illuminant C by adjusting the amount of light emitted by the red, green, and blue primaries in such a way that both halves of the test field are equal in their perceptual attributes of brightness, hue, and colorfulness. Performing this matching

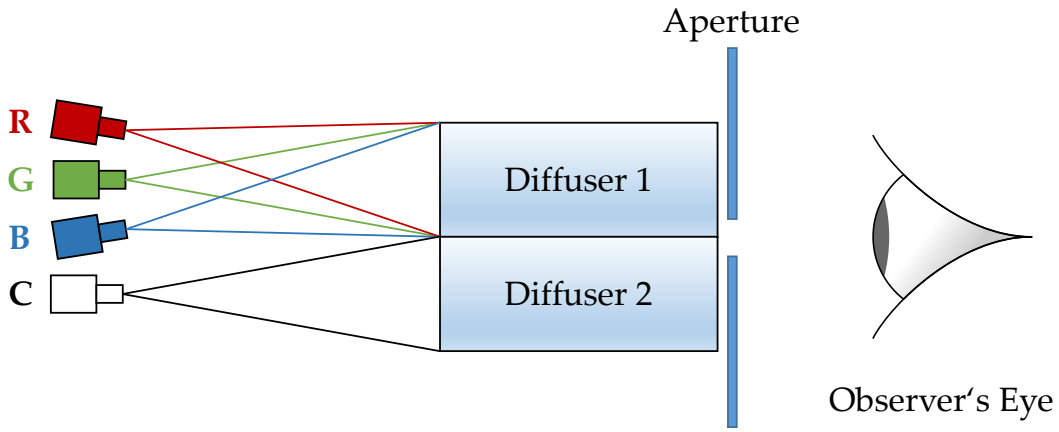


Figure 2.2 – Principle of trichromatic color matching based on additive mixing of lights according to the rules of Grassmann [70]. Light source C provides a monochromatic stimulus of a certain color impression to be matched by adjusting the amount of light emitted by the monochromatic red, green, and blue light sources defining the primaries of the color matching apparatus. The diffusers result in two uniform test fields illuminated by the respective light sources and presented side-by-side to be viewed and adjusted accordingly by the observer to match each other. Figure is reproduced from Ref. [68], p. 25.

process for each wavelength in the visible spectrum gives the ratios of the three primaries that are necessary to reproduce the perceived color of the reference stimulus for the whole spectral range. The resulting curves $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ are called color matching functions (CMFs) and are shown in Fig. 2.3 as obtained from the combined and averaged results of the trichromatic color matching experiments conducted by Wright [71] and Guild [72] upon which the corresponding CIE recommendation on colorimetry [69, 73] is based.

In both cases, the angular size of the matching field viewed by the observer was 2°. A total number of ten and seven individual observers were tested by Wright and Guild, respectively. Even though two distinct experimental setups using different methods to create the monochromatic stimuli had been applied, it was possible to combine the two sets of results by mathematically mapping the respective data onto what would have been obtained if the line spectra at 700 nm (red), 546.1 nm (green), and 435.8 nm (blue) had been used as the monochromatic matching primaries. Moreover, the original results were further transformed so that the amounts of red, green, and blue required to match the reference stimulus were not measured in units of luminance but in units defined with respect to the constraint that a perfect white of an equal-energy radiator (CIE illuminant E) should be matched by equal amounts of the three matching stimuli [68]. Based on this redefinition of units, the ordinate of Fig. 2.3 is defined.

As stated by Wright [71] and Guild [72], it was not feasible to reproduce every monochromatic reference stimulus of the visible part of the spectrum by simply mixing certain amounts of the matching primaries. Instead, there were certain wavelengths for which a small to moderate amount of one of the primaries had to be added to the reference stimulus to allow for a proper matching with the remaining two stimuli resulting in the negative parts of the CMFs. With the CMFs being known, they can be applied as weighting functions to determine the amounts R_p, G_p, and B_p of the primaries required to match an arbitrary test stimulus of know

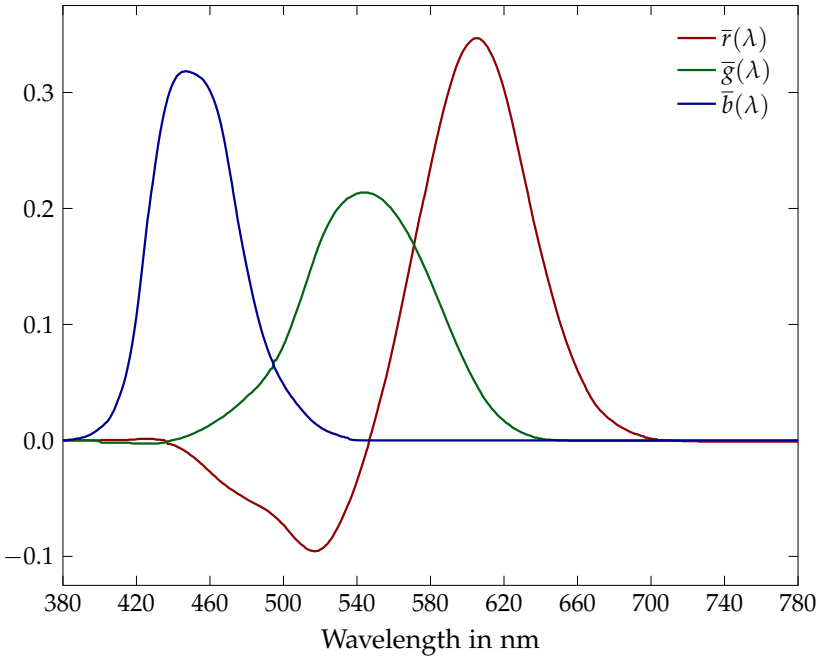


Figure 2.3 – Color matching functions as derived from the experiments of Wright [71] and Guild [72]. The three different curves represent the amounts of light of the red (700 nm), green (546.1 nm), and blue (435.8 nm) monochromatic primaries needed to match a monochromatic stimulus at a certain wavelength using transformed units such that equal amounts of them would match the color appearance of an equal-energy radiator.

spectral composition $S(\lambda)$. The amounts R_p, G_p, and B_p are called tristimulus values which due to additive color matching are given by

$$\begin{aligned} R_\mathrm{p} &= k \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{r}(\lambda) S(\lambda)\,\mathrm{d}\lambda, \\ G_\mathrm{p} &= k \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{g}(\lambda) S(\lambda)\,\mathrm{d}\lambda, \\ B_\mathrm{p} &= k \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{b}(\lambda) S(\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{2.1}$$

where k is a constant to scale the tristimulus values according to the radiometric unit in which $S(\lambda)$ is measured. If for example $S(\lambda)$ gives the spectral radiance of the test stimulus, the constant k is chosen such that

$$L = 1.0000 \cdot R_\mathrm{p} + 4.5907 \cdot G_\mathrm{p} + 0.0601 \cdot B_\mathrm{p}, \tag{2.2}$$

gives the respective luminance in $\mathrm{cd\,m^{-2}}$. If $S(\lambda)$ is measured in some other radiometric unit, L will give its photometric counterpart. Please note that the coefficients of Eq. (2.2) arise from the initial transformation into relative units constrained by the equal-energy radiator as described above.

Even though the tristimulus values R_p, G_p, and B_p would be sufficient to provide a colorimetric system allowing for precise color specification and the calculation of color properties from SPDs of arbitrary light sources, they were not adopted for defining CIE colorimetry. In

the early 1930s, when the CIE agreed upon a unified system of color specification, it was common understanding that the negative parts of the CMFs and, consequently, the fact that the tristimulus values of Eqs. (2.1) could exhibit negative numbers might cause problems in colorimetric calculations. Bearing in mind that computers had not been invented yet, this was for sure a legitimate concern.

Hence, it was decided to apply a linear transformation that converts from the system of real primaries to a system of imaginary primaries, where the resulting CMFs $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$ and, therefore, the corresponding tristimulus values X, Y, and Z only exhibit positive values. Further requirements were that the tristimulus values of an equal-energy radiator should still be equal, i.e., $X = Y = Z$, that one of the CMFs should represent the photopic spectral luminous efficiency function $V(\lambda)$ [74] offering the possibility of providing photometric quantities, and that the volume spanned by the new set of imaginary primaries should be as small as possible. The final form of the linear transformation for the tristimulus values is given by

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.768892 & 1.751748 & 1.130160 \\ 1.000000 & 4.590700 & 0.060100 \\ 0 & 0.056508 & 5.594292 \end{pmatrix} \begin{pmatrix} R_\mathrm{p} \\ G_\mathrm{p} \\ B_\mathrm{p} \end{pmatrix}. \tag{2.3}$$

It should be noted that the Y value sums up to the same photometric quantity as given by Eq. (2.2). Furthermore, with the CMFs basically representing the tristimulus values of monochromatic radiation, the same transformation matrix can be used to convert the $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, and $\overline{b}(\lambda)$ functions to the new $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$ CMFs which have finally been standardized by the CIE giving the CIE 1931 standard colorimetric observer shown in Fig. 2.4.

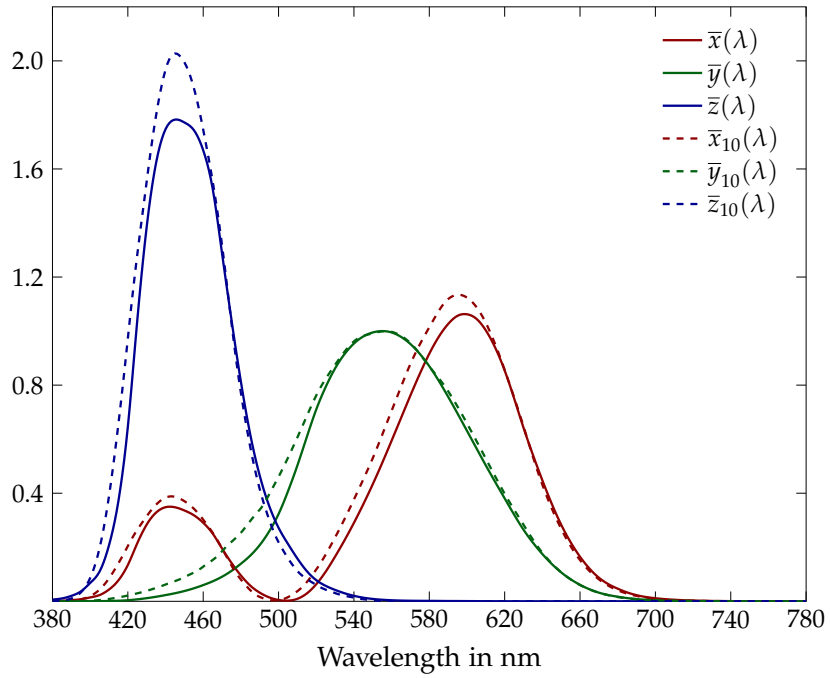


Figure 2.4 – Illustrations of CIE color matching functions of the 1931 and 1964 standard colorimetric observer represented by full and dashed lines, respectively.

As before, this new set of CMFs can be used as weighting functions to calculate the corresponding tristimulus values of an arbitrary color stimulus $\phi(\lambda)$ of light seen by the standardized observer. The resulting equations are

$$\begin{aligned} X &= K \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{x}(\lambda)\phi(\lambda)\,\mathrm{d}\lambda, \\ Y &= K \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{y}(\lambda)\phi(\lambda)\,\mathrm{d}\lambda, \\ Z &= K \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{z}(\lambda)\phi(\lambda)\,\mathrm{d}\lambda. \end{aligned} \tag{2.4}$$

In case of describing self-luminous objects, it is recommended to measure $\phi(\lambda)$ in radiometric units with K being set to the maximum of the luminous efficacy of radiation $K_\mathrm{m} = 683\,\mathrm{lm\,W^{-1}}$ [75] so that the resulting Y value represents the corresponding photometric quantity and the two remaining tristimulus values are scaled accordingly.

For non self-luminous objects which reflect or transmit light emitted from an external light source the situation is slightly different. Since only part of the light is reflected or transmitted by the object and eventually perceived by the observer, the color stimulus $\phi(\lambda)$ inserted in Eqs. (2.4) has to be modified as follows:

$$\phi(\lambda) = R(\lambda)S(\lambda) \quad \text{or} \quad \phi(\lambda) = T(\lambda)S(\lambda), \tag{2.5}$$

where $R(\lambda)$ and $T(\lambda)$ are the object's spectral reflectance and transmittance factor, respectively, and $S(\lambda)$ is the (relative) SPD of the external light source illuminating the object. In these cases, the constant K is chosen such that for a perfectly reflecting (white) or transmitting (translucent) object with $R(\lambda)$ or $T(\lambda)$ equal unity for all wavelengths, a maximum value of 100 is assigned to Y leading to

$$K = \frac{100}{\int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \bar{y}(\lambda)S(\lambda)\,\mathrm{d}\lambda}, \tag{2.6}$$

This definition basically accounts for the process of adaptation of the human visual system which ensures that a white surface is always seen in a white color independent of the amount of light falling onto it. Hence, for the human visual system it does not matter if a piece of white paper is observed in a relatively dim office environment or outdoors under direct sunlight: Its color appearance will always remain approximately constant. With this constancy in perception being also ascertainable for non-white objects [76, 77], it should be clear that a perfectly white reflecting target always represents the brightest object in an appreciated illumination scene so that the brightness of other colored objects being present can only be judged in relation to the brightness of that specific white object. In this sense, Eq. (2.6) is an arbitrary but very convenient choice for defining the constant K in case of considering non self-luminous objects.

So far, the colorimetric system standardized by the CIE in 1931 and discussed above was entirely based on color matching experiments examining only one viewing field size of 2°. However, in realistic applications the viewing angle under which colored objects are perceived by an observer is usually much larger. With the density and distribution of the cones on the retina varying considerably from one point to another [59, 78, 79], significant deviations of the CMFs are expected when the size of the viewing field is increased in a color

matching experiment. For this reason, similar experiments as those of Wright and Guild were repeated for the 10° case [80, 81]. Based on these investigations, the CIE finally derived a second set of transformed CMFs, $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$, and $\bar{z}_{10}(\lambda)$, also shown in Fig. 2.4, which were standardized in 1964 and denoted as the CIE 1964 standard colorimetric observer [73] recommended to be used for field sizes greater than 4°. Even though this new set of CMFs can be applied as weighting functions to obtain corresponding tristimulus values X_{10}, Y_{10}, and Z_{10} in the same manner as with the original ones of the 2° standard observer (see Eqs. (2.4)), an interpretation of the Y_{10} value in terms of photometric quantities for self-luminous objects is, in contrast to the CIE 1931 standard colorimetric system, not feasible. Instead, relative values as recommended for non self-luminous objects (see Eq. (2.6)) or any other convenient choice for the constant K_{10} shall be used.

2.2 COLOR REPRESENTATION AND UNIFORM COLOR SPACES

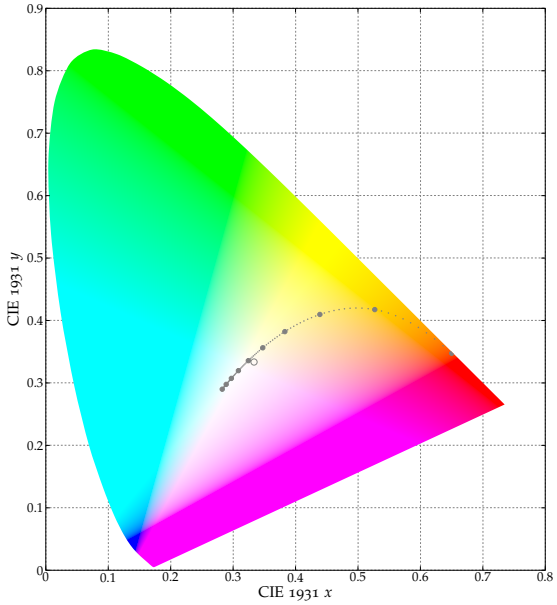
In many applications, it is useful to represent colors by their location in some kind of color space ideally separating the concept of color into its fundamental parts of brightness and chromaticity. An obvious choice of constructing such a three-dimensional color space would be the direct use of the tristimulus values. However, even though the Y value approximately relates to brightness perception, no such correlations with any of the perceptual color attributes can be observed for the X and Z values. Thus, other measures have to be derived from the tristimulus values to provide a proper representation of color specifications. As a first attempt, the CIE proposed to use relative magnitudes of the tristimulus values to describe the chromaticity of a stimulus. These so-called chromaticity coordinates are given by

$$\begin{aligned} x &= \frac{X}{X+Y+Z}, \\ y &= \frac{Y}{X+Y+Z}, \\ z &= \frac{Z}{X+Y+Z} = 1-x-y, \end{aligned} \tag{2.7}$$

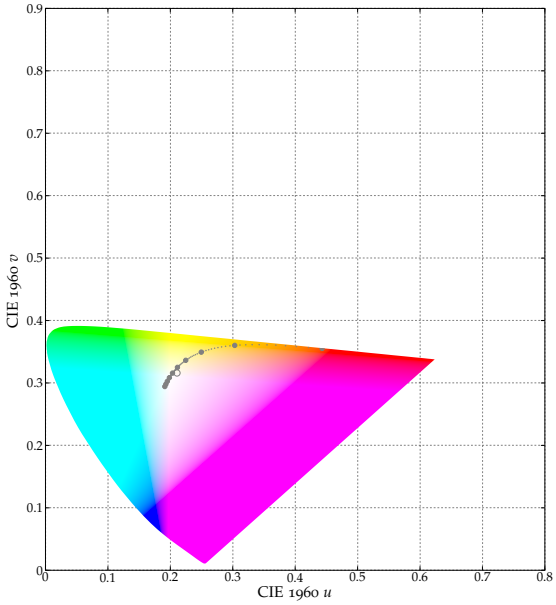
where due to redundancy in the formulae only the first two were needed to define what is today known as the CIE 1931 (x, y) chromaticity diagram. The application of this diagram, which is illustrated in Fig. 2.5(a), allows for a descriptive, two-dimensional representation of color specifications. Its horseshoe-shaped boundary is determined by the spectral locus as defined by the (x, y) coordinates of monochromatic stimuli and encloses all chromaticities that are theoretically perceivable by the standard human observer under a viewing angle of 2°. Of course, a similar chromaticity diagram can also be derived for the 10° observer.

Although such chromaticity diagrams can be helpful in order to get a first idea about the approximate nature of the stimulus to be perceived, it should be remembered that for an exact representation of human color perception several other important impact factors must be considered that have not been discussed yet. Besides the viewing angle under which the stimulus is appreciated, these are for example complexities such as the effect of the surrounding, simultaneous and successive contrasts as well as the observer's state of adaptation.

Another drawback of the simple chromaticity diagrams is their perceptual non-uniformity regarding the representation of color differences between two distinct color stimuli. Based on the work of MacAdam [82, 83], who examined the nature of visual sensitivity to small color differences, it could be shown that the size, shape, and orientation of ellipses representing



(a) Representation of the CIE 1931 (x, y) chromaticity diagram.



(b) Representation of the CIE 1960 (u, v) uniform chromaticity scale.

Figure 2.5 – Illustrations of the CIE 1931 and 1960 chromaticity diagrams for the 2° colorimetric standard observer. The outer curved boundaries are determined by the spectral loci as defined by the respective coordinates of monochromatic stimuli. They enclose all chromaticities given by the color shaded areas that are theoretically perceivable by the standard human observer. In addition, the locus of the Planckian radiator (dotted curve, filled circles) as well as the chromaticity of the equal-energy radiator (open circle) are indicated in both diagrams.

the observed standard deviations/tolerances of color matching vary strongly throughout the CIE 1931 chromaticity diagram in a manner which appeared to be systematic. However, if the chromaticity diagram was perceptually uniform, the tolerance boundaries should be represented by circles which all exhibit the same size no matter in which part of the diagram the color matching is performed. In other words, chromatictiy diagrams must be designed in such a way that approximately equal distances separate the points representing all pairs of equally bright colors that are just noticably different under certain specified viewing conditions.

In order to achieve such a property and to establish a first meaningful measure of color differences, Judd [84, 85] and MacAdam [86] himself investigated the possibility of constructing chromaticity diagrams of perceptual uniformity by transformationally distorting the original CIE 1931 (x, y) representation. The proposal of MacAdam, which reads

$$\begin{aligned} u &= \frac{4x}{12y - 2x + 3} = \frac{4X}{X + 15Y + 3Z}, \\ v &= \frac{6y}{12y - 2x + 3} = \frac{6Y}{X + 15Y + 3Z}, \end{aligned} \tag{2.8}$$

was later adopted by the CIE defining the CIE 1960 uniform chromaticity scale shown in Fig. 2.5(b). In a technical note, the CIE stated that MacAdam's proposal should be used *"whenever a diagram yielding color spacing perceptually more nearly uniform than the (x, y) diagram is desired"* [87]. Nowadays, the most prominent application of the CIE 1960 uniform chromaticity scale is the calculation of correlated color temperatures (CCTs) because it is the only accepted representation that exhibits isothermal lines being perpendicularly oriented to

the locus of the Planckian radiator which eventually allows for an accurate CCT computation (see Sec. 2.3). Furthermore, the color difference of two distinct equally bright color stimuli $i = 1, 2$, that at least to some extent resembles human perception, can easily be calculated by taking the Euclidean distance between their respective chromaticity coordinates (u_i, v_i). A further refinement regarding perceptual uniformity could be achieved by scaling Eqs. (2.8) as follows:

$$\begin{aligned} u' &= u, \\ v' &= \frac{3}{2}v. \end{aligned} \tag{2.9}$$

This new set of coordinates (u', v') basically defines another widely used chromaticity diagram know as the CIE 1976 uniform chromaticity scale.

So far, only two-dimensional chromaticity representations of colors have been discussed which are by definition only strictly applicable to colors that exhibit more or less the same brightness. However, as stated previously, colors in general differ in both chromaticity and brightness. Hence, in order to target the necessity of creating color representations including in the best case all aspects of human color vision, a vast variety of different three-dimensional color spaces have been devised and partly adopted by the CIE. For the sake of brevity, only those color spaces upon which the main findings and discussions of the current thesis are based should be discussed in the following sections.

2.2.1 *CIELAB Color Space and Color Difference Formulae*

One of the first further evolved color spaces, whose use was recommended by the CIE, is the so-called CIE 1976 $(L^*a^*b^*)$ color space commonly known as CIELAB. It is a mathematical description of all perceivable colors spanned by the lightness correlate L^* in combination with the two components a^* and b^* of the green-red and blue-yellow perception, respectively, defining an easy-to-interpret three-dimensional coordinate system, where, similar to a Cartesian system, all three axis are perpendicularly oriented to one another. Here, lightness can be conceived as the brightness of an object judged in relation to the brightness of a similarly illuminated perfect white. All CIELAB quantities can be derived from the tristimulus values X, Y, and Z as follows [73]:

$$\begin{aligned} L^* &= 116 f\left(\frac{Y}{Y_\mathrm{n}}\right) - 16, \\ a^* &= 500\left(f\left(\frac{X}{X_\mathrm{n}}\right) - f\left(\frac{Y}{Y_\mathrm{n}}\right)\right), \\ b^* &= 200\left(f\left(\frac{Y}{Y_\mathrm{n}}\right) - f\left(\frac{Z}{Z_\mathrm{n}}\right)\right), \end{aligned} \tag{2.10}$$

where

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{841}{108}t + \frac{4}{29} & \text{otherwise} \end{cases}, \tag{2.11}$$

with X_n, Y_n, and Z_n denoting the tristimulus values of an appropriately chosen reference white (usually the white point of the illumination under which the colors are observed).

From the above equations, correlates of hue h_{ab} and chroma C^*_{ab} can be derived which are given by

$$h_{ab} = \arctan\left(\frac{b^*}{a^*}\right), \tag{2.12}$$

and

$$C^*_{ab} = \sqrt{(a^*)^2 + (b^*)^2}. \tag{2.13}$$

Please note that, similar to the definition of lightness, chroma is defined as the colorfulness of a color stimulus judged in proportion to the brightness of a reference white and, therefore, can be considered as a measure of the relative colorfulness.

Due to the Cartesian-like nature of the CIELAB color space, it was an obvious choice to define the color difference between two stimuli $i = 1, 2$ to be given by their Euclidean distance

$$\begin{aligned} \Delta E^*_{ab} &= \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \\ &= \sqrt{\left(L^*_2 - L^*_1\right)^2 + \left(a^*_2 - a^*_1\right)^2 + \left(b^*_2 - b^*_1\right)^2}, \end{aligned} \tag{2.14}$$

which despite the beauty of its simplicity reveals one major drawback of the CIELAB color space as defined in 1976: It is still faraway from being perceptually uniform. Imagine for example that a just noticeable difference (JND) between two say red colors is assigned a certain ΔE^*_{ab}. If the assumption of perfect uniformity was fulfilled, the same value of ΔE^*_{ab} would be obtained for a JND in the blue (or any other) part of the color space. However, this is not the case for the CIELAB where significant variations in the ΔE^*_{ab} values of the JNDs or other defined color differences related to human color perception are observed throughout the color space [88–91].

Hence, in subsequent years a lot of effort mainly driven by the colorant industry was put into the attempt of defining more sophisticated color difference formulae based on the CIELAB coordinates intended to better correlate with human perception and to overcome the deficiencies of the color space. After some intermediate steps, some of which are still applied today (e.g., the CMC(l:c) formula devised by Clarke *et al.* [92]), the CIE finally came up in 2001 with the so-called CIEDE2000 formula [93], which is to date the latest CIE recommendation of calculating small color differences ($\Delta E^*_{ab} \lesssim 5$) [73]. Originally developed by a collaboration of several members of the CIE Technical Committee 1-47, the methodology of deriving the respective forumlae in combination with a step-by-step guide of the calculation process were first published by Luo *et al.* [94, 95]. As stressed by these authors, the new CIEDE2000 approach includes not only lightness, chroma, and hue weighting functions but also a hue-rotation term to allow for an interaction between chroma and hue differences for improving its predictive performance in the blue region of color space as well as a coordinate scaling factor to enhance the performance for colors close to the achromatic axis. Assuming a pair of color values in CIELAB color space (L^*_1, a^*_1, b^*_1) and (L^*_2, a^*_2, b^*_2), the respective color difference ΔE_{00} is then evaluated as follows:

STEP 1: Calculate modified chroma C'_i and hue h'_i values:

$$
\begin{aligned}
C^*_{ab,i} &= \sqrt{\left(a^*_i\right)^2 + \left(b^*_i\right)^2}, \\
\overline{C}^*_{ab} &= \frac{C^*_{ab,1} + C^*_{ab,2}}{2}, \\
G &= 0.5 \cdot \left(1 - \sqrt{\frac{\left(\overline{C}^*_{ab}\right)^7}{\left(\overline{C}^*_{ab}\right)^7 + (25)^7}} \right), \\
a'_i &= a^*_i \left(1 + G\right), \\
C'_i &= \sqrt{\left(a'_i\right)^2 + \left(b^*_i\right)^2}, \\
h'_i &= \begin{cases} 0 & \text{if } a'_i = b^*_i = 0 \\ \tan^{-1}\left(\frac{b^*_i}{a'_i}\right) & \text{otherwise} \end{cases},
\end{aligned}
\tag{2.15}
$$

where $i = 1, 2$ denotes the reference and test color sample, respectively.

STEP 2: Calculate differences of the CIELAB lightness and the modified chroma and hue correlates indicated by $\Delta L'$, $\Delta C'$, and $\Delta H'$, respectively:

$$
\begin{aligned}
\Delta L' &= L^*_2 - L^*_1, \\
\Delta C' &= C'_2 - C'_1, \\
\Delta h' &= \begin{cases} 0 & \text{if } C'_1 C'_2 = 0 \\ (h'_2 - h'_1) - 360^\circ & \text{if } C'_1 C'_2 \neq 0 \text{ and } (h'_2 - h'_1) > 180^\circ \\ (h'_2 - h'_1) + 360^\circ & \text{if } C'_1 C'_2 \neq 0 \text{ and } (h'_2 - h'_1) < -180^\circ \\ h'_2 - h'_1 & \text{otherwise} \end{cases}, \\
\Delta H' &= 2\sqrt{C'_1 C'_2} \sin\left(\frac{\Delta h'}{2}\right).
\end{aligned}
\tag{2.16}
$$

STEP 3: Calculate the weighting functions S_L, S_C, and S_H:

$$\begin{aligned}
\overline{L}' &= \frac{L_1^* + L_2^*}{2}, \\
\overline{C}' &= \frac{C_1' + C_2'}{2}, \\
\overline{h}' &= \begin{cases} h_1' + h_2' & \text{if } C_1'C_2' = 0 \\ \frac{h_1'+h_2'+360^\circ}{2} & \text{if } C_1'C_2' \neq 0 \text{ and } |h_1' - h_2'| > 180^\circ \text{ and } h_1' + h_2' < 360^\circ \\ \frac{h_1'+h_2'-360^\circ}{2} & \text{if } C_1'C_2' \neq 0 \text{ and } |h_1' - h_2'| > 180^\circ \text{ and } h_1' + h_2' \geq 360^\circ \\ \frac{h_1'+h_2'}{2} & \text{otherwise} \end{cases}, \\
T &= 1 - 0.17\cos\left(\overline{h}' - 30^\circ\right) + 0.24\cos\left(2\overline{h}'\right) + 0.32\cos\left(3\overline{h}' + 6^\circ\right) - 0.2\cos\left(4\overline{h}' - 63^\circ\right), \\
S_L &= 1 + \frac{0.015\left(\overline{L}' - 50\right)^2}{\sqrt{20 + \left(\overline{L}' - 50\right)^2}}, \\
S_C &= 1 + 0.045\overline{C}', \\
S_H &= 1 + 0.015\overline{C}'T.
\end{aligned} \tag{2.17}$$

STEP 4: Combine everything to obtain the CIEDE2000 color difference ΔE_{00}:

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)}, \tag{2.18}$$

where

$$R_T = -\sin\left(2\Delta\theta\right) R_C, \tag{2.19}$$

with

$$\begin{aligned}
\Delta\theta &= 30 \cdot \exp\left(-\left(\frac{\overline{h}' - 275^\circ}{25}\right)^2\right), \\
R_C &= 2\sqrt{\frac{\left(\overline{C}'\right)^7}{\left(\overline{C}'\right)^7 + (25)^7}},
\end{aligned} \tag{2.20}$$

represents the rotation parameter to correct for the distortions of the CIELAB color space in the blue region where, in contrast to other parts of the hue circle, the main axes of perceptual tolerance ellipses do not point towards the origin of the color space [68, 88]. Hence, a hue-dependent rotational transformation of the color coordinates is performed and incorporated into Eq. (2.18) in order to achieve the desired hue linearity [96]. Furthermore, the weighted lightness, chroma, and hue components are additionally corrected by some parametric factors k_L, k_C, and k_H, respectively. These parameters serve to adjust the ΔE_{00} formula

according to the characteristics of the individual color samples under inspection (e.g., texture, surroundings, sample separation etc.). Nevertheless, setting all three parameters to unity has proven to be a good choice for most applications [68] and should be pursued throughout this thesis. However, in some specific industrial applications where an accurate and proper mold validation of sample pairs plays a crucial role for the production process (e.g., in the coating or textile industry), an empirical adjustment of these parameters based on accepted/rejected color samples might be indispensable [96].

2.2.2 *Uniform Color Spaces Based on the CIECAM02 Color-appearance Model*

Even though the CIEDE2000 approach significantly improved the predictive performance of the CIELAB color space with regard to perceptual color differences [94, 98–100], which at the same time gives strong confidence that the ΔE_{00} formula is reliable [101, 102], its derivation is still based on a color space that is neither perceptually uniform nor in perfect agreement with human color vision in the sense that a more sophisticated chromatic adaptation transform is not incorporated. As pointed out by Hunt [68], the CIELAB system was originally intended for being applied to color samples assessed under adaptation conditions not too different from daylight. Hence, it would be desirable to develop a model of color vision that can be used for predicting the appearance of object colors for a wide range of different viewing conditions. Such color-appearance models (CAMs) have intensively been discussed in the literature [103–119].

Based on these discussions, the CIE in 2004 eventually approved a CAM called CIECAM02 to become the recommended standard in color appearance evaluation [120–122], which provides measures for related colors (i.e., reflecting samples observed in non-isolated viewing

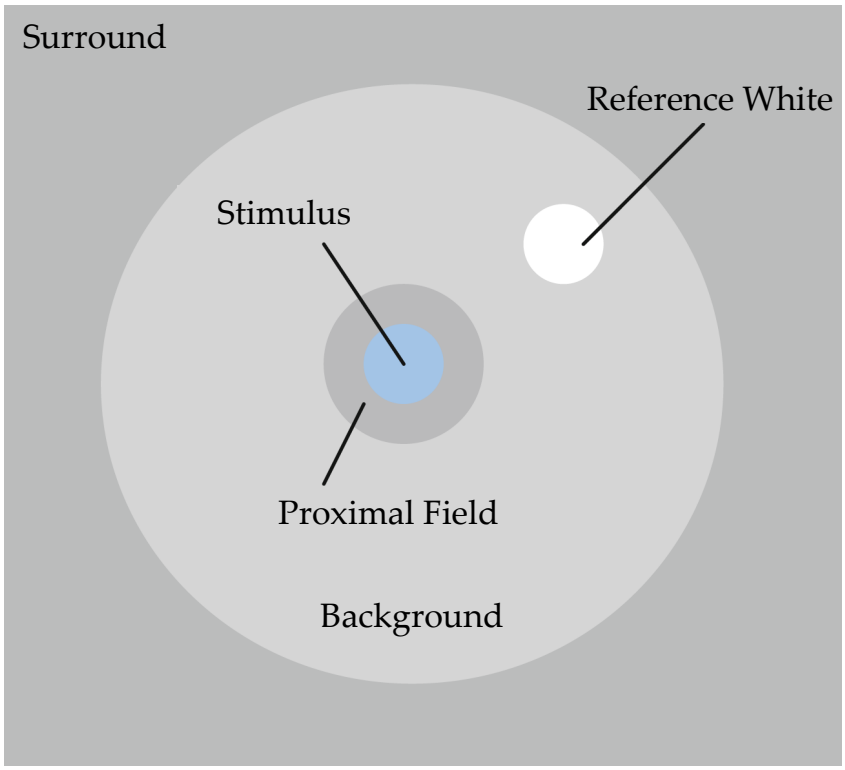


Figure 2.6 – Standardized viewing configuration of related colors in the CIECAM02 color-appearance model. Figure is reproduced from Ref. [97], p. 22.

conditions under white light) that do not only correlate with lightness, hue, and chroma, as offered in the CIELAB system, but also with the remaining perceptual absolute and relative color attributes of brightness, colorfulness, and saturation. Here, the latter is understood as the colorfulness of an area judged in relation to its perceived brightness.

In order to simplify a usually complex viewing scene to make it suitable for modeling, the CIECAM02 approach assumes five different visual fields shown in Fig. 2.6 to include the most important factors that have an impact on the color appearance of related colors. These are:

- The actual color stimulus to be assessed. Typically a uniform object/color patch of about 2° angular dimension.
- The proximal field which represents the immediate surrounding of the color stimulus showing a typical angular extension of about 2° measured from the edge of the stimulus in all or most directions.
- The background which denotes the surrounding of the color stimulus exhibiting an angular extension of typically 10° from the edge of the proximal field in all or most directions. Often, the background is of the same color and luminance as the proximal field so that the former can be regarded as extending directly from the edge of the stimulus.
- The surround which represents the general surrounding field outside the background.
- The adapting field which is the total environment in which a color stimulus is perceived including the proximal field, the background, and the surround. It basically extends to the limits of vision in all directions.

Please note that for the following considerations and also for the discussions provided later in this thesis, a visual conflation of the proximal field and the background is assumed so that both are considered to exhibit the same color and luminance. Hence, based on the viewing configuration described above, the following steps published in Ref. [120] have to be taken in order to calculate from the assessment of a color sample under test illumination the various perceptual correlates provided by the CIECAM02 model:

STEP 1: Select the appropriate surround conditions from the evaluation of the surround ratio S_R given by

$$S_R = \frac{L_{SW}}{L_{DW}}, \tag{2.21}$$

where L_{SW} is the luminance of the surround white and L_{DW} is the luminance of the device or scene white (adopted white point) under the test illuminant both measured in $\mathrm{cd\,m^{-2}}$. If S_R equals zero, a dark surround should be chosen. If S_R is smaller than 0.2, a dim surround applies. And, last but not least, an S_R equal or larger than 0.2 corresponds to an average surround. The respective surround parameters can be extracted from Table 2.1, where F is the parameter of adaptation, N_c is the chromatic surround induction factor having an impact on the perceived chroma, and c denotes the exponential non-linearity factor used for the calculation of the correlates of both lightness and brightness.

Table 2.1 – Model parameters of the categorical surround conditions in CIECAM02.

Surround Condition	c	N_c	F
Average	0.69	1.0	1.0
Dim	0.59	0.9	0.9
Dark	0.525	0.8	0.8

STEP 2: Calculate background and adapting field parameters

$$\begin{aligned} n &= \frac{Y_\mathrm{b}}{Y_\mathrm{w}}, \\ N_\mathrm{bb} = N_\mathrm{cb} &= 0.725 \cdot \left(\frac{1}{n}\right)^{0.2}, \\ z &= 1.48 + \sqrt{n}, \\ k &= \frac{1}{5L_\mathrm{A} + 1}, \\ F_\mathrm{L} &= 0.2 \cdot k^4 \left(5L_\mathrm{A}\right) + 0.1 \cdot \left(1 - k^4\right)^2 \left(5L_\mathrm{A}\right)^{\frac{1}{3}}, \end{aligned} \tag{2.22}$$

where Y_w is the Y tristimulus value of the device or scene white (adopted white) and L_A is the luminance of the adapting field which is usually set to $1/5$ of the luminance of the adopted white. Furthermore, Y_b is the background luminous factor ($\widehat{=}$ relative luminance of the background) defined by

$$Y_\mathrm{b} = 100 \cdot \frac{L_\mathrm{b}}{L_\mathrm{DW}}, \tag{2.23}$$

where L_b is the background luminance measured in $\mathrm{cd\,m^{-2}}$. In most applications Y_b can be set to 20 [97].

STEP 3: Convert the relative tristimulus values of the color sample X, Y, Z and the test illuminant X_w, Y_w, Z_w to sharpened spectral sensitivity responses by using the CAT02 chromatic adaptation transform matrix:

$$\begin{aligned} \begin{pmatrix} R \\ G \\ B \end{pmatrix} &= M_\mathrm{CAT02} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \\ \begin{pmatrix} R_\mathrm{w} \\ G_\mathrm{w} \\ B_\mathrm{w} \end{pmatrix} &= M_\mathrm{CAT02} \begin{pmatrix} X_\mathrm{w} \\ Y_\mathrm{w} \\ Z_\mathrm{w} \end{pmatrix}, \end{aligned} \tag{2.24}$$

with

$$M_\mathrm{CAT02} = \begin{pmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{pmatrix}. \tag{2.25}$$

STEP 4: Compute the D factor which determines the degree of adaptation to the scene white point under the test illuminant. If complete adaptation can be assumed, D should be set to unity.

$$D = F\left(1 - \frac{1}{3.6} \cdot \exp\left(\frac{-L_A - 42}{92}\right)\right). \tag{2.26}$$

STEP 5: Perform the weighted chromatic adaptation including the D factor to obtain the R, G, B responses for the corresponding colors under CIE reference illuminant E denoted by a subscript c:

$$\begin{aligned}
R_c &= \left(\frac{Y_w}{R_w}D + 1 - D\right) R, \\
G_c &= \left(\frac{Y_w}{G_w}D + 1 - D\right) G, \\
B_c &= \left(\frac{Y_w}{B_w}D + 1 - D\right) B, \\
R_{w,c} &= \left(\frac{Y_w}{R_w}D + 1 - D\right) R_w, \\
G_{w,c} &= \left(\frac{Y_w}{G_w}D + 1 - D\right) G_w, \\
B_{w,c} &= \left(\frac{Y_w}{B_w}D + 1 - D\right) B_w.
\end{aligned} \tag{2.27}$$

STEP 6: Convert to Hunt-Pointer-Estévez (HPE) cone response space

$$\begin{aligned}
\begin{pmatrix} \rho \\ \gamma \\ \beta \end{pmatrix} &= M_{HPE} M_{CAT02}^{-1} \begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix}, \\
\begin{pmatrix} \rho_w \\ \gamma_w \\ \beta_w \end{pmatrix} &= M_{HPE} M_{CAT02}^{-1} \begin{pmatrix} R_{w,c} \\ G_{w,c} \\ B_{w,c} \end{pmatrix},
\end{aligned} \tag{2.28}$$

where the transformation matrices are given by

$$M_{CAT02}^{-1} = \begin{pmatrix} 1.0961 & -0.278869 & 0.182745 \\ 0.454369 & 0.473533 & 0.072098 \\ -0.009628 & -0.005698 & 1.015326 \end{pmatrix}, \tag{2.29}$$

and

$$M_{HPE} = \begin{pmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.00000 & 0.00000 & 1.00000 \end{pmatrix}. \tag{2.30}$$

STEP 7: Apply luminance-level adaptation and non-linear compression

$$
\begin{aligned}
\rho_{\mathrm{a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \rho}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \rho}{100} \right)^{0.42}} \right], \\
\gamma_{\mathrm{a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \gamma}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \gamma}{100} \right)^{0.42}} \right], \\
\beta_{\mathrm{a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \beta}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \beta}{100} \right)^{0.42}} \right], \\
\rho_{\mathrm{w,a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \rho_{\mathrm{w}}}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \rho_{\mathrm{w}}}{100} \right)^{0.42}} \right], \\
\gamma_{\mathrm{w,a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \gamma_{\mathrm{w}}}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \gamma_{\mathrm{w}}}{100} \right)^{0.42}} \right], \\
\beta_{\mathrm{w,a}} &= 0.1 + \left[\frac{400 \left(\frac{F_{\mathrm{L}} \beta_{\mathrm{w}}}{100} \right)^{0.42}}{27.13 + \left(\frac{F_{\mathrm{L}} \beta_{\mathrm{w}}}{100} \right)^{0.42}} \right].
\end{aligned}
\tag{2.31}
$$

Please note that in case that ρ, γ, β, ρ_{w}, γ_{w}, or β_{w} are negative, their absolute values must be used and the expression in brackets $[\,\cdot\,]$ must be made negative.

STEP 8: Calculate color-difference signals a_{02} and b_{02} and hue angle h

$$
\begin{aligned}
a_{02} &= \rho_{\mathrm{a}} - \frac{12\gamma_{\mathrm{a}}}{11} + \frac{\beta_{\mathrm{a}}}{11}, \\
b_{02} &= \frac{1}{9} \cdot \left(\rho_{\mathrm{a}} + \gamma_{\mathrm{a}} - 2\beta_{\mathrm{a}} \right), \\
h &= \tan^{-1} \left(\frac{b_{02}}{a_{02}} \right).
\end{aligned}
\tag{2.32}
$$

Please note that h must be converted to degrees.

STEP 9: Calculate the eccentricity factor e_{t} and the hue quadrature H:

$$
\begin{aligned}
e_{\mathrm{t}} &= \frac{1}{4} \cdot \left(\cos \left(h \frac{\pi}{180} + 2 \right) + 3.8 \right), \\
H &= H_i + \frac{100 \cdot \left(h' - h_i \right) / e_i}{\left(h' - h_i \right) / e_i + \left(h_{i+1} - h' \right) / e_{i+1}},
\end{aligned}
\tag{2.33}
$$

Table 2.2 – Unique hue data for calculating the hue quadrature.

	Red	Yellow	Green	Blue	Red
i	1	2	3	4	5
h_i	20.14°	90.00°	164.25°	237.53°	380.14°
e_i	0.8	0.7	1.0	1.2	0.8
H_i	0.0	100.0	200.0	300.0	400.0

where the unique hue data shown in Table 2.2 should be used to perform the linear interpolation for calculating the quadrature. Please note that an appropriate value of i must be chosen so that $h_i \leq h' < h_{i+1}$ with

$$h' = \begin{cases} h + 360° & \text{if } h < h_1 \\ h & \text{otherwise} \end{cases}. \tag{2.34}$$

STEP 10: Calculate the achromatic responses of the stimulus, A^*, and of the scene white, A^*_{w}:

$$\begin{aligned} A^* &= \left(2\rho_{\mathrm{a}} + \gamma_{\mathrm{a}} + \frac{1}{20}\cdot\beta_{\mathrm{a}} - 0.305\right) N_{\mathrm{bb}}, \\ A^*_{\mathrm{w}} &= \left(2\rho_{\mathrm{w,a}} + \gamma_{\mathrm{w,a}} + \frac{1}{20}\cdot\beta_{\mathrm{w,a}} - 0.305\right) N_{\mathrm{bb}} \end{aligned} \tag{2.35}$$

STEP 11: Calculate the perceptual correlates of lightness, J, brightness, Q, chroma, C, colorfulness, M, and saturation, s:

$$\begin{aligned} t &= \frac{(50000/13)\, N_c N_{\mathrm{cb}} \cdot e_{\mathrm{t}} \sqrt{a_{02}^2 + b_{02}^2}}{\rho_{\mathrm{a}} + \gamma_{\mathrm{a}} + \frac{21}{20}\cdot\beta_{\mathrm{a}}}, \\ J &= 100 \cdot \left(\frac{A^*}{A^*_{\mathrm{w}}}\right)^{cz}, \\ Q &= \frac{4}{c}\cdot\sqrt{\frac{J}{100}}\,(A^*_{\mathrm{w}} + 4)\, F_{\mathrm{L}}^{0.25}, \\ C &= t^{0.9}\cdot\sqrt{\frac{J}{100}}\,(1.64 - 0.29^n)^{0.73}, \\ M &= C \cdot F_{\mathrm{L}}^{0.25}, \\ s &= 100 \cdot \sqrt{\frac{M}{Q}}. \end{aligned} \tag{2.36}$$

As a summary of the bunch of complex equations given above, it should be emphasized that the CIECAM02 model is specifically designed to transform the input tristimulus values of a color sample perceived under a wide range of more or less arbitrary viewing conditions to the corresponding correlates of perceptual color attributes as viewed under adaptation to

the white point of a reference illuminant, which in case of CIECAM02 is given by the CIE illuminant E (equal-energy radiator). If the color stimulus has an angular dimension of more than 4°, it is recommended to use the CIE 1964 standard colorimetric observer to calculate the input tristimulus values [68], otherwise the CIE 1931 standard colorimetric observer should be used.

Hence, it can be concluded that the CIECAM02 model provides so far the most complete description of human color vision enabling the accurate prediction of the perceptional color appearance of object colors. Due to the implementation of a sophisticated chromatic adaptation transform and the additional consideration of the surrounding parameters, even cross-comparisons of color stimuli assessed under different adaptation and viewing conditions become feasible. This might be useful when comparing for example the appearance of a real object/model/item that should be portrayed for a magazine with its eventually obtained photographic reproduction on a calibrated computer monitor. Ideally, these two color appearances of the same object should be matched despite being perceived under different viewing conditions. By applying CIECAM02 such an adjustment can more or less easily be performed, making it the recommended and commonly used choice for establishing an accurate and reliable color management system [120, 123–125], which besides including a proper chromatic adaptation transform also has the power to model some of the well known perceptual phenomena affecting color appearance. These are the Hunt [126], Stevens [127], Surround [128], and Lightness-Contrast [129] effects which are all implicitly embodied in the CIECAM02 formulae.

As proposed by Luo *et al.* [130], the CIECAM02 model further allows for the derivation of an associated three-dimensional color representation space. They stated that an appropriate color space for universal use, i.e., for being applied to small, moderate, and large color differences, is spanned by the three coordinates

$$\begin{aligned} a'_{\mathrm{M}} &= M' \cos(h)\,, \\ b'_{\mathrm{M}} &= M' \sin(h)\,, \\ J' &= 1.7 \cdot \frac{J}{1 + 0.007 \cdot J}, \end{aligned} \tag{2.37}$$

with

$$M' = \frac{1}{0.0228} \cdot \ln(1 + 0.0228 \cdot M)\,, \tag{2.38}$$

where J, M, and h are the CIECAM02 correlates of lightness, colorfulness, and hue angle, respectively. The corresponding color difference between two arbitrary color samples $i = 1, 2$ is then computed by

$$\begin{aligned} \Delta E' &= \sqrt{(\Delta J')^2 + \left(\Delta a'_{\mathrm{M}}\right)^2 + \left(\Delta b'_{\mathrm{M}}\right)^2} \\ &= \sqrt{\left(J'_2 - J'_1\right)^2 + \left(a'_{\mathrm{M},2} - a'_{\mathrm{M},1}\right)^2 + \left(b'_{\mathrm{M},2} - b'_{\mathrm{M},1}\right)^2}. \end{aligned} \tag{2.39}$$

It could be shown that the latter equation provides good to excellent correlations with various data sets of perceptual color differences typically used for model performance testing [97, 130]. In this context, the CIECAM02-UCS outperforms most of the available color space or color difference formula alternatives. Furthermore, the CIECAM02-UCS offers magnificent perceptual uniformity as can be concluded from the analysis of the characteristics of

experimentally determined color discrimination ellipses which, in case of the CIECAM02-UCS, resemble circles of approximately equal sizes for all hue regions as shown by Luo *et al.* [97, 130]. Since it provides excellent model performance based on the features of human color vision, the CIECAM02-UCS in conjunction with its color difference formula given by Eq. (2.39) will be adopt for the main part of this thesis.

2.3 COMPUTATION OF CORRELATED COLOR TEMPERATURE

Last but not least, a short introduction to the calculation of correlated color temperatures (CCTs) should be given. The proper determination of the CCT of an arbitrary white light source plays an important role in many lighting applications. Besides providing a tool for categorizing the perceived white of an illuminant (e.g., warm white, neutral, cool white), the CIE recommendation for specifying the color rendering properties of white light sources, as it will be discussed in the next chapter, is based on the concept of CCT. From a colorimetric point of view, the CCT of a light source represents the temperature a Planckian radiator of the same brightness would have whose perceived color most closely conforms the perceived color of that light source.

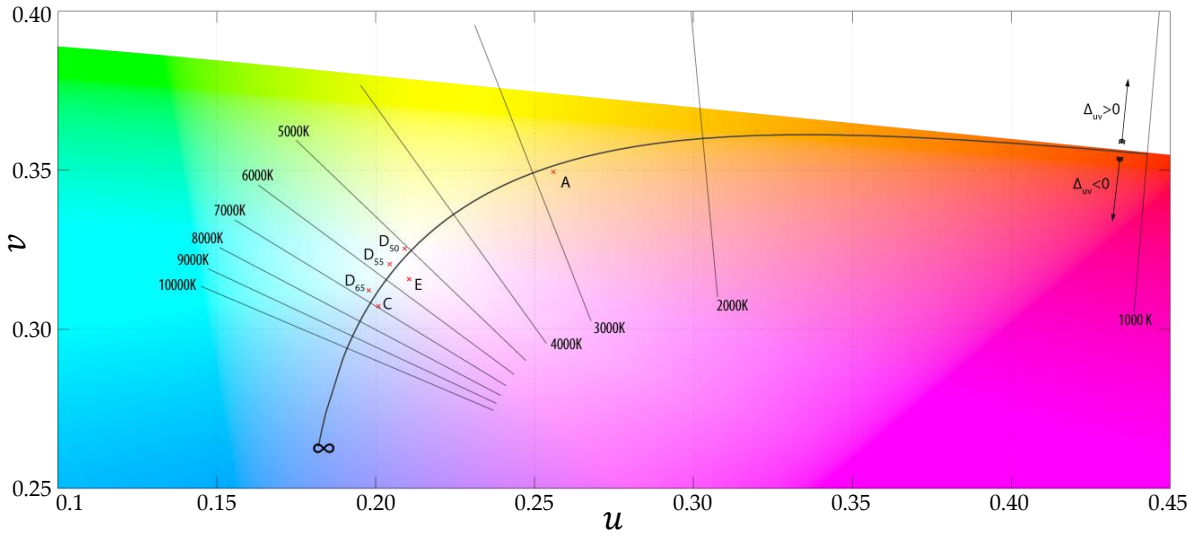


Figure 2.7 – Illustration of the Planckian locus on the CIE 1960 (u, v) chromaticity diagram. The normals on the Planckian loucs are isothermal lines originally defined by Judd [84] indicating equal correlated color temperature. Figure is reproduced from Ref. [131]

As mentioned previously, the CIE 1960 (u, v) uniform chromaticity scale is used to calculate CCTs. In this chromaticity diagram, the isothermal lines shown in Fig. 2.7 are perpendicularly oriented to the locus of the Planckian radiator. Hence, the CCT calculation for a given stimulus with (u_k, v_k) can be performed by simply determining the temperature corresponding to the point on the Planckian locus that is closest to (u_k, v_k). However, the only accurate way to do this is by solving a non-trivial nonlinear optimization problem. In the past, several different studies were therfore published trying to estimate the CCT of a light source by using simplifying methods and/or assumptions [132–140]. However, for the sake of accuracy, a new approach devised by Li *et al.* [141] should be applied here.

Their proposal is based on Newton's method of optimization [142] which makes use of the first and second order derivatives of the objective function $f(T)$ intended to be minimized. In the case of the CCT calculation, $f(T)$ is defined by the square of the distance between the

chromaticities (u_k, v_k) of the stimulus and the chromaticities $(u_P(T), v_P(T))$ of the Planckian radiator of given temperature T. In their paper, Li *et al.* derived analytically exact expressions for $f(T)$ and its corresponding first and second order derivatives allowing for an accurate determination of the temperature T that minimizes the objective function by applying Newton's method. It could further be shown that this newly proposed method outperforms all available alternatives cited above and, therefore, represents the method of choice for achieving highly accurate CCT estimates. At this point, it is important to note that since Newton's method by definition only converges locally, a good initial guess is necessary for convergence. As stated by Li *et al.*, their proposed method works outstandingly well for CCTs ranging from 500 K to 10^6 K with an initial guess provided by a method based on an explicit interpolation formula for isothermal line graphs which was originally developed by Robertson [132] and, over the years, widely used in practice [68]. Due to its excellent performance, this combination of an initial guess provided by the Robertson formula and a subsequent optimization as proposed Li *et al.* was finally chosen for the various CCT calculations performed throughout this thesis. Finally, it should noted that the concept of CCT is only defined for white light sources exhibiting a maximum distance of $\Delta uv = 5.0 \cdot 10^{-2}$ to the Planckian locus in CIE 1960 uv color space [73].

REFERENCES OF CHAPTER II

[57] J. D. Mollon, "Color Vision," *Annual Review of Psychology*, vol. 33, pp. 41–85, 1982.

[58] J. D. Mollon and J. K. Bowmaker, "The Spatial Arrangement of Cones in the Primate Fovea," *Nature*, vol. 360, no. 6405, pp. 38–44, 1992.

[59] H. Hofer, J. Carroll, J. Neitz, *et al.*, "Organization of the Human Trichromatic Cone Mosaic," *Journal of Neuroscience*, vol. 25, no. 42, pp. 9669–9679, 2005.

[60] D. Williams, "Color and the Cone Mosaic," in *Proceedings of the 14th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2006, pp. 1–2.

[61] D. H. Brainard, "Color and the Cone Mosaic," *Annual Review of Vision Science*, vol. 1, pp. 519–546, 2015.

[62] A. Stockman and L. T. Sharpe, "The Spectral Sensitivities of the Middle- and Long-Wavelength-Sensitive Cones Derived from Measurements in Observers of known Genotype," *Vision Research*, vol. 40, no. 13, pp. 1711–1737, 2000.

[63] A. Stockman, L. T. Sharpe, and C. Fach, "The Spectral Sensitivity of the Human Short-wavelength Sensitive Cones Derived from Thresholds and Color Matches," *Vision Research*, vol. 39, no. 17, pp. 2901–2927, 1999.

[64] Commission Internationale de l'Éclairage, "Fundamental Chromaticity Diagram with Physiological Axes – Part 1," *CIE Technical Report 170-1:2006*, 2006.

[65] H. J. A. Dartnall, J. K. Bowmaker, and J. D. Mollon, "Human Visual Pigments: Microspectrophotometric Results from the Eyes of Seven Persons," *Proceedings of the Royal Society of London B*, vol. 220, no. 1218, pp. 115–130, 1983.

[66] J. K. Bowmaker and H. J. A. Dartnall, "Visual Pigments of Rods and Cones in a Human Retina," *Journal of Physiology*, vol. 298, no. 1, pp. 501–511, 1980.

[67] P. K. Brown and G. Wald, "Visual Pigments in Single Rods and Cones of the Human Retina," *Science*, vol. 144, no. 3614, pp. 45–46 & 51–52, 1964.

[68] R. W. G. Hunt and M. R. Pointer, *Measuring Colour*, 4th. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2011.

[69] J. Schanda, "CIE Colorimetry," in *Colorimetry: Understanding the CIE System*, J. Schanda, Ed., CIE Preprint Edition. Vienna: CIE Central Bureau, 2006, ch. 3, pp. 1–54.

[70] H. G. Grassmann, "Zur Theorie der Farbenmischung," *Annalen der Physik*, vol. 165, no. 5, pp. 69–84, 1853.

[71] W. D. Wright, "A Re-determination of the Trichromatic Coefficients of the Spectral Colours," *Transactions of the Optical Society*, vol. 30, no. 4, pp. 141–164, 1929.

[72] J. Guild, "The Colorimetric Properties of the Spectrum," *Philosophical Transactions of the Royal Society of London A*, vol. 230, pp. 149–187, 1931.

[73] Commission Internationale de l'Éclairage, *Colorimetry*, CIE Technical Report 15:2004, 3rd ed. 2004.

[74] Commission Internationale de l'Éclairage, "The Basis of Physical Photometry," *CIE Technical Report 18.2*, 1983.

[75] G. Wiszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd, Wiley Classics Library Edition. New York, NY, USA: John Wiley & Sons, Inc., 2000.

[76] D. Jameson and L. M. Hurvich, "Essay Concerning Color Constancy," *Annual Review of Psychology*, vol. 40, pp. 1–22, 1989.

[77] D. H. Foster, "Color Constancy," *Vision Research*, vol. 51, no. 7, pp. 674–700, 2011.

[78] A. Roorda and D. R. Williams, "The Arrangement of the Three Cone Classes in the Living Human Eye," *Nature*, vol. 397, no. 6719, pp. 520–522, 1999.

[79] J. Carroll, J. Neitz, and M. Neitz, "Estimates of L:M Cone Ratio from ERG Flicker Photometry and Genetics," *Journal of Vision*, vol. 2, no. 8, pp. 531–542, 2002.

[80] W. S. Stiles and J. Burch, "N.P.L. Colour-matching Investigation: Final Report," *Optica Acta*, vol. 6, no. 1, pp. 1–26, 1959.

[81] N. I. Speranskaya, "Determination of Spectrum Color Coordinates for 27 Normal Observers," *Optics and Spectroscopy*, vol. 7, pp. 424–428, 1959.

[82] D. L. MacAdam, "Visual Sensitivities to Color Differences in Daylight," *Journal of the Optical Society of America*, vol. 32, no. 5, pp. 247–274, 1942.

[83] D. L. MacAdam, "Specification of Small Chromaticity Differences," *Journal of the Optical Society of America*, vol. 33, no. 1, pp. 18–26, 1943.

[84] D. B. Judd, "A Maxwell Triangle Yielding Uniform Chromaticity Scales," *Journal of the Optical Society of America*, vol. 25, no. 1, pp. 24–35, 1935.

[85] D. B. Judd, "Estimation of Chromaticity Differences and Nearest Color Temperature on the Standard 1931 I. C. I. Colorimetric Coordinate System," *Journal of the Optical Society of America*, vol. 26, no. 11, pp. 421–426, 1936.

[86] D. L. MacAdam, "Projective Transformations of I. C. I. Color Specifications," *Journal of the Optical Society of America*, vol. 27, no. 8, pp. 294–299, 1937.

[87] Commission Internationale de l'Éclairage, "Technical Note: Brussels Session of the International Commission on Illumination," *Journal of the Optical Society of America*, vol. 50, no. 1, pp. 89–90, 1960.

[88] A. R. Robertson, "The CIE 1976 Color-Difference Formulae," *Color Research & Application*, vol. 2, no. 1, pp. 7–11, 1977.

[89] S. M. Jaekel, "Utility of Color-difference Formulas for Match-acceptability Decisions," *Applied Optics*, vol. 12, no. 6, pp. 1299–1316, 1973.

[90] M. R. Pointer, "A Comparison of the CIE 1976 Colour Spaces," *Color Research & Application*, vol. 6, no. 2, pp. 108–118, 1981.

[91] D. I. Morley, R. Munn, and F. W. Billmeyer, "Small and Moderate Colour Differences: II. The Morley Data," *Journal of the Society of Dyers and Colourists*, vol. 91, no. 7, pp. 229–242, 1975.

[92] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC79 Colour–difference Formula," *Journal of the Society of Dyers and Colourists*, vol. 100, no. 4, pp. 128–132, 1984.

[93] Commission Internationale de l'Éclairage, "Improvement to Industrial Colour Difference Evaluation," *CIE Technical Report 142:2001*, 2001.

[94] M. R. Luo, G. Cui, and B. Rigg, "The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000," *Color Research & Application*, vol. 26, no. 5, pp. 340–350, 2001.

[95] G. Sharma, W. Wencheng, and E. N. Dalal, "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.

[96] E. Rohner and D. C. Rich, "Eine Angenähert Gleichförmige Metrik für Industrielle Farbtoleranzen von Körperfarben," *Die Farbe*, vol. 42, no. 4, pp. 207–220, 1996.

[97] M. R. Luo and C. Li, "CIECAM02 and Its Recent Developments," in *Advanced Color Image Processing and Analysis*, C. Fernandez-Maloigne, Ed. New York, USA: Springer Science+Business Media, 2013, pp. 19–58.

[98] M. Melgosa, R. Huertas, and R. S. Berns, "Relative Significance of the Terms in the CIEDE2000 and CIE94 Color-difference Formulas," *Journal of the Optical Society of America A*, vol. 21, no. 12, pp. 2269–2275, 2004.

[99] M. Melgosa, "Improvement of CMC upon CIEDE2000 for a New Experimental Dataset," *Color Research & Application*, vol. 31, no. 3, pp. 239–241, 2006.

[100] S. Shen and R. S. Berns, "Color-difference Formula Performance for Several Datasets of Small Color Differences Based on Visual Uncertainty," *Color Research & Application*, vol. 36, no. 1, pp. 15–26, 2011.

[101] G. Cui, M. R. Luo, B. Rigg, *et al.*, "Colour-Difference Evaluation Using CRT Colours. Part I: Data Gathering and Testing Colour Difference Formulae," *Color Research & Application*, vol. 26, no. 5, pp. 394–402, 2001.

[102] M. R. Luo, "Development of Colour-difference Formulae," *Review of Progress in Coloration and Related Topics*, vol. 32, no. 1, pp. 28–39, 2002.

[103] T. Seim and A. Valberg, "Towards a Uniform Color Space: A Better Formula to Describe the Munsell and OSA Color Scales," *Color Research & Application*, vol. 11, no. 1, pp. 11–24, 1986.

[104] Y. Nayatani, K. Takahama, and H. Sobagaki, "Prediction of Color Appearance under Various Adaptation Conditions," *Color Research & Application*, vol. 11, no. 1, pp. 62–71, 1986.

[105] Y. Nayatani, K. Hashimoto, K. Takahama, *et al.*, "Whiteness-blackness and Brightness Response in a Non-linear Color-appearance Model," *Color Research & Application*, vol. 12, no. 3, pp. 121–127, 1987.

[106] Y. Nayatani, K. Hashimoto, K. Takahama, *et al.*, "A Non-linear Color-appearance Model Using Estévez-Hunt-Pointer Primaries," *Color Research & Application*, vol. 12, no. 5, pp. 231–242, 1987.

[107] Y. Nayatani, K. Takahama, H. Sobagaki, *et al.*, "Color Appearance Model and Chromatic Adaptation Transform," *Color Research & Application*, vol. 15, no. 4, pp. 210–221, 1990.

[108] Y. Nayatani, H. Sobagaki, K. Hashimoto, *et al.*, "Field Trials of a Non-linear Color-appearance Model," *Color Research & Application*, vol. 22, no. 4, pp. 240–258, 1997.

[109] M. D. Fairchild and R. S. Berns, "Image Color-appearance specification through Extension of CIELAB," *Color Research & Application*, vol. 18, no. 3, pp. 178–190, 1993.

[110] M. D. Fairchild, "Refinement of the RLAB Color Space," *Color Research & Application*, vol. 21, no. 5, pp. 338–346, 1996.

[111] R. W. G. Hunt, "Perceptual Factors Affecting Colour Order Systems," *Color Research & Application*, vol. 10, no. 1, pp. 12–19, 1985.

[112] R. W. G. Hunt, "A Model for Colour Vision for Predicting Colour Appearance," *Color Research & Application*, vol. 7, no. 2, pp. 95–112, 1982.

[113] R. W. G. Hunt and M. R. Pointer, "A Colour-appearance Transform for the CIE 1931 Standard Colorimetric Observer," *Color Research & Application*, vol. 10, no. 3, pp. 165–179, 1985.

[114] R. W. G. Hunt, "A Visual Model for Predicting Colour Appearance under Various Viewing Conditions," *Color Research & Application*, vol. 12, no. 6, pp. 297–314, 1987.

[115] R. W. G. Hunt, "Hue Shifts in Unrelated and Related Colours," *Color Research & Application*, vol. 14, no. 5, pp. 235–239, 1989.

[116] R. W. G. Hunt, "Revised Colour-appearance Model for Related and Unrelated Colours," *Color Research & Application*, vol. 16, no. 3, pp. 146–165, 1991.

[117] R. W. G. Hunt, "An Improved Predictor of Colourfulness in a Model of Colour Vision," *Color Research & Application*, vol. 19, no. 1, pp. 23–26, 1994.

[118] R. W. G. Hunt and M. R. Luo, "Evaluation of a Model of Colour Vision by Magnitude Scalings: Discussion of Collected Results," *Color Research & Application*, vol. 19, no. 1, pp. 27–33, 1994.

[119] M. R. Luo, M.-C. Lo, and W.-G. Kuo, "The LLAB(*l*:*c*) Colour Model," *Color Research & Application*, vol. 21, no. 6, pp. 412–429, 1996.

[120] Commission Internationale de l'Éclairage, "A Colour-appearance Model for Colour Management Systems: CIECAM02," *CIE Technical Report 159:2004*, 2004.

[121] N. Moroney, M. D. Fairchild, R. G. W. Hunt, *et al.*, "The CIECAM02 Color-appearance Model," in *Proceedings of the 10th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2002, pp. 23–27.

[122] C. Li, M. R. Luo, R. G. W. Hunt, *et al.*, "The Performance of CIECAM02," in *Proceedings of the 10th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2002, pp. 28–32.

[123] "ICC Profiles, Color Appearance Modeling, and the Microsoft Windows Color System," in *Color Management: Understanding and using ICC Profiles*, P. Green and M. A. Kriss, Eds. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2010, pp. 53–55.

[124] I. Tastl, M. Bhachech, N. Moroney, *et al.*, "ICC Colour Management and CIECAM02," in *Proceedings of the 13th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2005, pp. 217–223.

[125] R. Gury and M. Shaw, "Dealing with Imaginary Color Encodings in CIECAM02 in an ICC workflow," in *Proceedings of the 13th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2005, pp. 318–318.

[126] R. W. G. Hunt, "Light and Dark Adaptation and Perception of Color," *Journal of the Optical Society of America*, vol. 42, no. 3, pp. 190–199, 1952.

[127] J. C. Stevens and S. S. Stevens, "Brightness Functions: Effects of Adaptation," *Journal of the Optical Society of America*, vol. 53, no. 3, pp. 375–385, 1963.

[128] C. J. Bartleson and E. J. Breneman, "Brightness Perception in Complex Fields," *Journal of the Optical Society of America*, vol. 57, no. 7, pp. 953–957, 1967.

[129] M. R. Luo, X. W. Gao, and S. A. R. Sciviner, "Quantifying Colour Appearance. Part V. Simultaneous Contrast," *Color Research & Application*, vol. 20, no. 1, pp. 18–28, 1995.

[130] M. R. Luo, G. Cui, and C. Li, "Uniform Colour Spaces Based on CIECAM02 Colour Appearance Model," *Color Research & Application*, vol. 31, no. 4, pp. 320–330, 2006.

[131] Wikipedia User Adoniscik, *Planckian locus in the CIE* 1960 *UCS using the Judd-Vos* 2° *CMF*, `https://en.wikipedia.org/wiki/CIE_1960_color_space`, [Online; accessed 16-April-2018], 2008.

[132] A. R. Robertson, "Computation of Correlated Color Temperature and Distribution Temperature," *Journal of the Optical Society of America*, vol. 58, no. 11, pp. 1528–1535, 1968.

[133] J. Schanda and M. Dányi, "Correlated Color Temperature Calculations in the CIE 1976 Chromaticity Diagram," *Color Research & Application*, vol. 2, no. 4, pp. 161–163, 1977.

[134] J. Schanda, M. Mészáros, and G. Czibula, "Calculating Correlated Color Temperature with a Desktop Programmable Calculator," *Color Research & Application*, vol. 3, no. 2, pp. 65–69, 1978.

[135] M. Krystek, "An Algorithm to Calculate Correlated Colour Temperature," *Color Research & Application*, vol. 10, no. 1, pp. 38–40, 1985.

[136] J. L. Gardner, "Correlated Colour Temperature Uncertainty and Estimation," *Metrologia*, vol. 37, no. 5, pp. 381–384, 2000.

[137] X. Qiu, "Formulas for Computing Correlated Color Temperature," *Color Research & Application*, vol. 12, no. 5, pp. 285–287, 1987.

[138] C. S. McCamy, "Correlated Color Temperature as an Explicit Function of Chromaticity Coordinates," *Color Research & Application*, vol. 17, no. 2, pp. 142–144, 1992.

[139] C. S. McCamy, "Correlated Color Temperature as an Explicit Function of Chromaticity Coordinates (Erratum)," *Color Research & Application*, vol. 18, no. 2, p. 150, 1993.

[140] J. Hernández-Andrés, R. L. Lee, and J. Romero, "Calculating Correlated Color Temperatures Across the Entire Gamut of Daylight and Skylight Chromaticities," *Applied Optics*, vol. 38, no. 27, pp. 5703–5709, 1999.

[141] C. Li, G. Cui, M. Melgosa, *et al.*, "Accurate Method for Computing Correlated Color Temperature," *Optics Express*, vol. 24, no. 13, pp. 14 066–14 078, 2016.

[142] B. T. Polyak, "Newton's Method and its Use in Optimization," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1086–1096, 2007.

[143] D. Kaye, *BrainyQuote.com: Quotes of Danny Kaye*, `https://www.brainyquote.com/quotes/danny_kaye_125475`, [Online; accessed 22-May-2018].

Part 3

MEMORY AND PREFERRED COLORS IN COLOR RENDERING

"Life is a great big canvas; you should throw all the paint you can on it"

– Danny Kaye [143]

3

MEMORY AND PREFERRED COLORS IN COLOR RENDERING

With the CIE colorimetry discussed in the previous chapter providing a framework to model human color vision, the question of how this framework must be applied to assess the color quality of an arbitrary white light source arises. As stated in the introduction of this thesis, the advent of fluorescent gas-discharge lamps and later of light emitting diodes (LEDs) introduced a completely new spectral flexibility in designing light emitting devices which made developers and manufacturers strive after new tools and algorithms allowing for tweaking the light emission of an illuminant towards high user acceptability and preference. In general, preference in lighting applications is always a matter of how appealing the objects in an illuminated scenery appear to the individual observer, which is a very subjective aspect of color quality making a comprehensive description of the associated phenomena a very challenging task. On the other hand, it would be favorable to have a universally valid, single metric which can be used in a large variety of different situations to easily appraise whether the light emitted by an illuminant is judged to be of high or poor quality when assessed by an observer. In the past, the CIE as well as a large number of other researchers all over the globe addressed this topic – often with rather moderate success.

In this context, the most promising approaches in specifying the perceived color quality of light sources in real lighting applications are those based on internal references such as preferred or memory colors. The reason for this is that in the absence of an external reference, i.e., a reference illuminant, daylight, etc., which is usually the case for the majority of lighting applications, object colors and, consequently, the color quality of an illumination are always judged in relation to what people expect the perceived objects to look like. Depending on the specific application, this could be either a question of preference, of naturalness, or of any other subjective aspect of color quality. In any case, it always involves the comparison of the perceived scene with an inherent reference which generally has to be recalled from memory.

The following sections should therefore give a short introduction of how memory and preferred colors have been used so far in the literature to specify color rendering properties of white light sources in order to provide the basis for the further considerations discussed in the proceeding chapters of this thesis. As a reference method, the standard CIE procedure will be presented first. Afterwards, the focus will be solely on the specifically memory- and/or preference-based approaches.

3.1 COLOR RENDERING OF WHITE LIGHT SOURCES

In order to evaluate and quantify the ability of artificial light sources to render object colors with respect to certain aspects of color quality such as fidelity, preference, naturalness, vividness, and attractiveness [23] several different color quality metrics have been proposed in the literature [16, 19, 23, 26, 27, 30, 144–161]. Even though some of these candidates provide quite good correlations with the more subjective aspects of color perception like preference and attractiveness, the most widely used method in industry and application is still the so-called color rendering index (CRI) R_a originally introduced in 1965 [16] and updated twice in 1974 [158] and 1995 [159] by the CIE, which is purely based on a fidelity evaluation. In principle, the CRI mathematically compares the color appearance of eight different color samples

(see Fig. 3.1) illuminated by a test light source to the color appearance the same test samples would show when being illuminated by a reference light source. By definition, a value of $R_a = 100$ is attributed to the reference light source and the maximum value a test light source could achieve, in case that the calculated color differences of the test samples under reference and test illuminant are negligibly small.

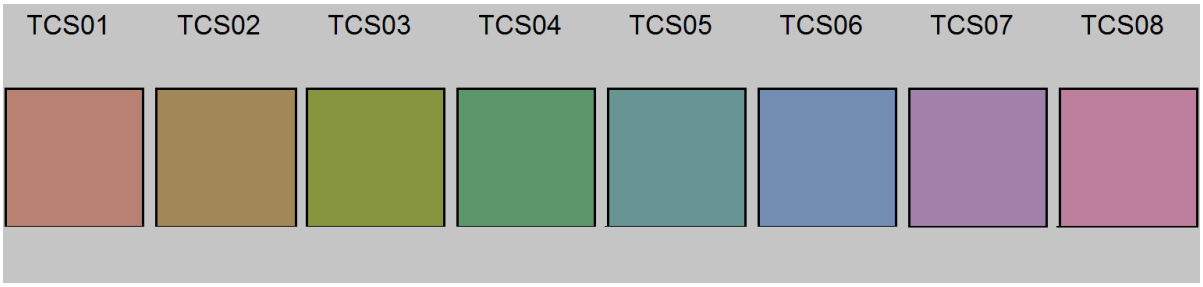


Figure 3.1 – Overview of the eight Munsell test color samples used for evaluating the CIE color rendering index R_a. The colors with their corresponding Munsell notation given in parenthesis are from left to right: light grayish red (7.5 R 6/4), dark grayish yellow (5 Y 6/4), strong yellow green (5 GY 6/8), moderate yellowish green (2.5 G 6/6), light bluish green (10 BG 6/4), light blue (5 PB 6/8), light violet (2.5 P 6/8), and light reddish purple (10 P 6/8).

In Fig. 3.2, the corresponding calculation scheme is illustrated. Being based on a test color method, the CIE 1931 tristimulus values of the various test color samples (TCSs) shown in Fig. 3.1 must be determined first for both the test and reference light source. With the original TCSs being taken from an early edition of the Munsell color atlas, it is hard to find nowadays physically accurate copies of these samples. For this reason, the CIE specified a set of eight spectral reflectances, representing the original TCSs, which should be used for calculating the CRI from a suitably accurate spectroradiometric measurement of the illuminant to be tested. These spectral reflectances can be extracted from the latest CIE publication on measuring and specifying color rendering properties of light sources [159] and shall be used together with the emitted spectra of the test and reference light source to calculate the corresponding tristimulus values following Eqs. (2.4).

According to the CIE recommendation, the reference light source must show nearly the same chromaticity as the test illuminant. For this reason, a selection criteria based on the CCT of the test light source was proposed. For CCTs below 5000 K a Planckian radiator of the form

$$M(\lambda, T_{cp}) = \frac{c_1}{\lambda^5} \cdot \frac{1}{\exp(\frac{c_2}{\lambda \cdot T_{cp}}) - 1}, \tag{3.1}$$

shall be used, where $M(\lambda, T)$ is the spectral radiant exitance of a black body with absolute temperature T in Kelvin, $c_1 = 3.7418 \cdot 10^{-16}\,\mathrm{W\,m^2}$ is the first radiation constant, $c_2 = 1.4388 \cdot 10^{-2}\,\mathrm{K\,m}$ is the second radiation constant, and T_{cp} is the CCT of the test light source. For CCT values larger than 5000 K a corresponding daylight phase shall be used as reference whose relative SPD is defined by

$$S(\lambda) = S_0(\lambda) + M_1 S_1(\lambda) + M_2 S_2(\lambda), \tag{3.2}$$

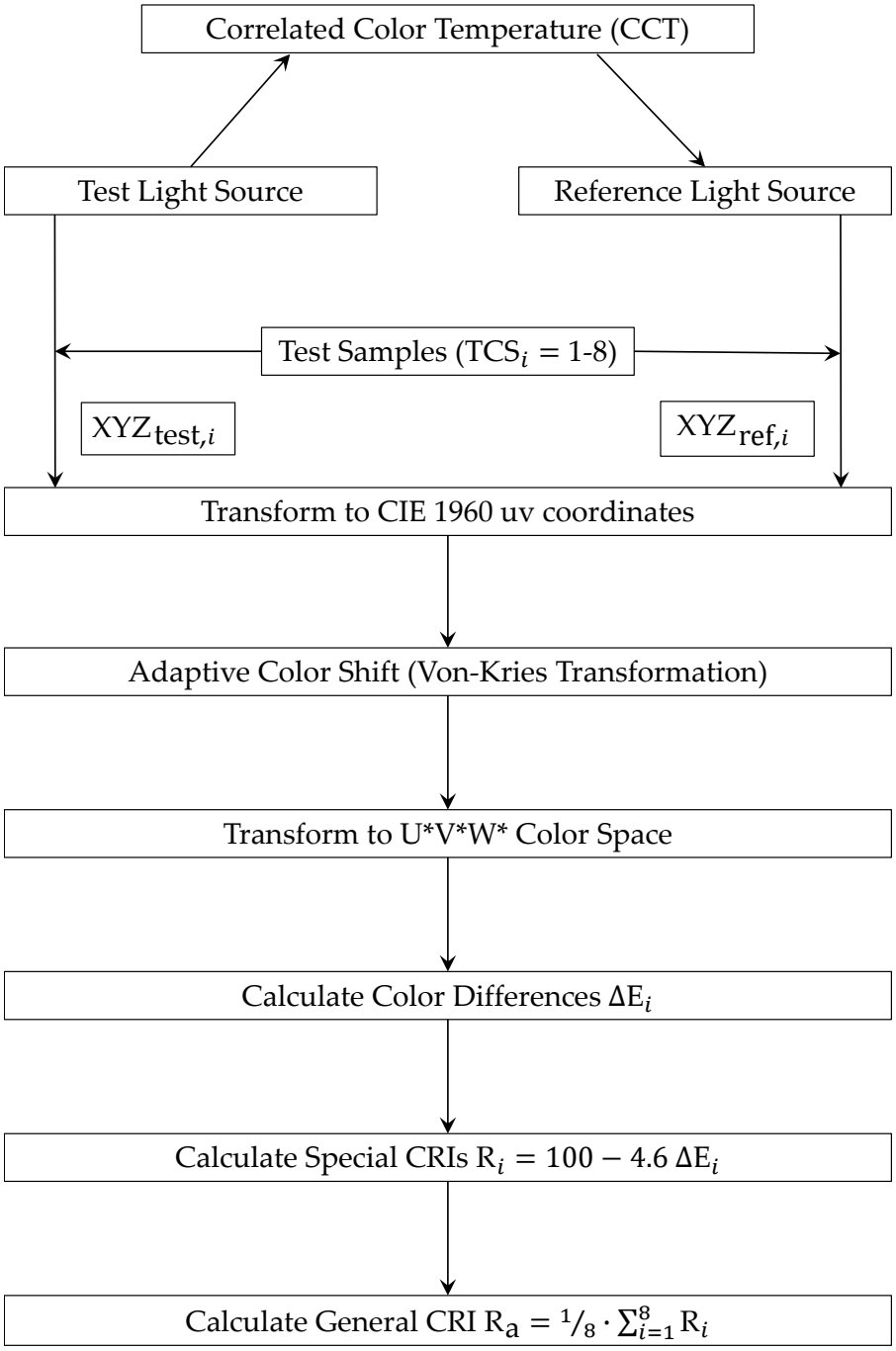


Figure 3.2 – Calculation scheme of the CIE color rendering index R_a. The von-Kries transformation to account for the adaptive color shift due to the different state of chromatic adaptation under test and reference light source is performed using CIE 1960 *uv* coordinates, whereas for the comparison of the color appearance of the eight test color samples illuminated by these two light sources the CIE 1964 $U^*V^*W^*$ uniform space was adopted.

where the functions $S_0(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$ are tabulated in the latest CIE publication on colorimetry [73] and the factors M_1 and M_2 are given by

$$M_1 = \frac{-1.3515 - 1.7703x_D + 5.9114y_D}{0.0241 + 0.2562x_D - 0.7341y_D},$$
$$M_2 = \frac{0.03000 - 31.4424x_D + 30.0717y_D}{0.0241 + 0.2562x_D - 0.7341y_D}. \tag{3.3}$$

Here, x_D and y_D are the respective CIE 1931 chromaticity coordinates. For CCTs up to 7000 K the former is given by

$$x_\mathrm{D} = \frac{-4.6070 \cdot 10^9}{T_\mathrm{cp}^3} + \frac{2.9678 \cdot 10^6}{T_\mathrm{cp}^2} + \frac{0.09911 \cdot 10^3}{T_\mathrm{cp}} + 0.244063, \tag{3.4}$$

whereas for CCTs larger than 7000 K the slightly modified expression

$$x_\mathrm{D} = \frac{-2.0064 \cdot 10^9}{T_\mathrm{cp}^3} + \frac{1.9018 \cdot 10^6}{T_\mathrm{cp}^2} + \frac{0.24748 \cdot 10^3}{T_\mathrm{cp}} + 0.237040 \tag{3.5}$$

should be used. In both cases, the latter reads

$$y_\mathrm{D} = -3.000 x_\mathrm{D}^2 + 2.870 x_\mathrm{D} - 0.275, \tag{3.6}$$

which basically constraints the locus of the daylight phases for any given T_cp. Please note again that the concept of CCT and, therefore, a meaningful CRI calculation is only defined for white light sources exhibiting a maximum distance of $\Delta uv = 5.0 \cdot 10^{-2}$ to the Planckian locus in CIE 1960 uv color space [73]. However, the CIE additionally stated that with differences between test and reference illuminant larger than $\Delta uv = 5.4 \cdot 10^{-3}$ care must be taken for interpretation since the resulting CRI may be expected to become less accurate [159].

Once the reference illuminant is determined, the tristimulus values of the TCSs under both test and reference light source can be calculated. Next, a transformation to the perceptionally more uniform CIE 1960 color space is performed. The corresponding color coordinates of the i^th TCS are given by

$$\begin{aligned} u_{\mathrm{k,r},i} &= \frac{4X_{\mathrm{k,r},i}}{X_{\mathrm{k,r},i} + 15Y_{\mathrm{k,r},i} + 3Z_{\mathrm{k,r},i}}, \\ v_{\mathrm{k,r},i} &= \frac{6Y_{\mathrm{k,r},i}}{X_{\mathrm{k,r},i} + 15Y_{\mathrm{k,r},i} + 3Z_{\mathrm{k,r},i}}, \end{aligned} \tag{3.7}$$

where the indices $\cdot_\mathrm{k}$ and $\cdot_\mathrm{r}$ designate the coordinates of the test and reference light source, respectively. In addition to the color coordinates $(u_{\mathrm{k},i}, v_{\mathrm{k},i})$ and $(u_{\mathrm{r},i}, v_{\mathrm{r},i})$ of the $i = 1, ..., 8$ Munsell color samples rendered under test and reference illuminant, the color coordinates $(u_\mathrm{k}, v_\mathrm{k})$ and $(u_\mathrm{r}, v_\mathrm{r})$ of the respective white points must also be calculated.

In order to account for the adaptive color shift due to the different state of chromatic adaptation regarding the two different illumination white points, a von-Kries like adaptation transformation is applied which reads [159]

$$\begin{aligned} u'_{\mathrm{k},i} &= \frac{10.872 + 0.404\frac{c_\mathrm{r}}{c_\mathrm{k}} c_{\mathrm{k},i} - 4\frac{d_\mathrm{r}}{d_\mathrm{k}} d_{\mathrm{k},i}}{16.518 + 1.481\frac{c_\mathrm{r}}{c_\mathrm{k}} c_{\mathrm{k},i} - \frac{d_\mathrm{r}}{d_\mathrm{k}} d_{\mathrm{k},i}}, \\ v'_{\mathrm{k},i} &= \frac{5.520}{16.518 + 1.481\frac{c_\mathrm{r}}{c_\mathrm{k}} c_{\mathrm{k},i} - \frac{d_\mathrm{r}}{d_\mathrm{k}} d_{\mathrm{k},i}}, \end{aligned} \tag{3.8}$$

where the functions

$$\begin{aligned} c_m &= \frac{1}{v_m}\left(4 - u_m - 10 v_m\right), \\ d_m &= \frac{1}{v_m}\left(1.708 v_m + 0.404 - 1.481 u_m\right), \end{aligned} \tag{3.9}$$

shall be evaluated for the white points $(u_\mathrm{k}, v_\mathrm{k})$ and $(u_\mathrm{r}, v_\mathrm{r})$ of the test and reference light source giving $(c_\mathrm{k}, d_\mathrm{k})$ and $(c_\mathrm{r}, d_\mathrm{r})$, respectively, as well as for the coordinates $(u_{\mathrm{k},i}, v_{\mathrm{k},i})$ of the i^{th} TCS illuminated by the test light source giving $(c_{\mathrm{k},i}, d_{\mathrm{k},i})$ in the above equations. Please note that after performing the chromatic adaptation, the white point of the test light source has virtually been shifted onto the white point of the reference, i.e., $u'_\mathrm{k} = u_\mathrm{r}$ and $v'_\mathrm{k} = v_\mathrm{r}$.

Hence, the subsequently required transformation to the CIE 1964 color space originally proposed by Wyszecki [162] can be performed using the following equations for the TCS rendered by the test and reference illuminant, respectively:

$$\begin{aligned} W^*_{\mathrm{k},i} &= 25\left(Y_{\mathrm{k},i}\right)^{\frac{1}{3}} - 17, \\ U^*_{\mathrm{k},i} &= 13 W^*_{\mathrm{k},i}\left(u'_{\mathrm{k},i} - u_\mathrm{r}\right), \\ V^*_{\mathrm{k},i} &= 13 W^*_{\mathrm{k},i}\left(v'_{\mathrm{k},i} - v_\mathrm{r}\right), \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} W^*_{\mathrm{r},i} &= 25\left(Y_{\mathrm{r},i}\right)^{\frac{1}{3}} - 17, \\ U^*_{\mathrm{r},i} &= 13 W^*_{\mathrm{r},i}\left(u_{\mathrm{r},i} - u_\mathrm{r}\right), \\ V^*_{\mathrm{r},i} &= 13 W^*_{\mathrm{r},i}\left(v_{\mathrm{r},i} - v_\mathrm{r}\right). \end{aligned} \tag{3.11}$$

Finally, for each TCS the color difference of its color appearance as perceived under the test and reference illuminant is calculated by taking the Euclidean distance

$$\Delta E_i = \sqrt{\left(U^*_{\mathrm{r},i} - U^*_{\mathrm{k},i}\right)^2 + \left(V^*_{\mathrm{r},i} - V^*_{\mathrm{k},i}\right)^2 + \left(W^*_{\mathrm{r},i} - W^*_{\mathrm{k},i}\right)^2}. \tag{3.12}$$

The special CRIs are then defined as

$$R_i = 100 - 4.6\Delta E_i, \tag{3.13}$$

while the general CRI is given as their arithmetical mean

$$R_\mathrm{a} = \frac{1}{8}\sum_{i=1}^{8} R_i. \tag{3.14}$$

As can be seen, the special CRIs R_i are scaled in such a way that i) a value of 100 indicates a perfect agreement in the color appearance of the i^{th} TCS when the performances of both light sources are compared and ii) a general CRI of 51 is assigned to the standard F4 warm white fluorescent lamp which is basically achieved by the factor 4.6.

Even though the CRI R_a is still the standard recommendation of the CIE for assessing the color rendering properties of a light source being widely used in industry and application, limitations of such a measure, especially when it comes to perception-related color quality effects beyond color fidelity and/or when white LED light sources are involved, have extensively been reported in the literature [163–173] and various approaches to overcome these limitations have been proposed [19, 23, 26, 27, 30, 144–157, 160, 161].

Regarding the development of an improved fidelity index, the CIE, after long years of struggle, finally came up in 2017 with its new approach of a general color fidelity index R_f providing a scientifically accurate measure of color rendering with respect to a certain reference illuminant [152]. Similar to the CRI, it compares the color appearance of a set of TCSs illuminated by a test light source to the color appearance these test samples would show

when being illuminated by the reference light source. However, instead of using an outdated color space and chromatic adaptation transform, the R_f metric models the color appearance of the TCSs perceived under both light sources by adopting the more sophisticated and better suited CIECAM02 color appearance model, as introduced in the previous chapter, in conjunction with the CIE 1964 standard colorimetric observer (10° CMFs, see Fig. 2.4). In this framework, color differences between test and reference illumination are calculated in terms of CIECAM02-UCS color coordinates and a more adequate method of choosing the proper reference illuminant especially in the range between 4000 K and 5000 K has been introduced.

It is commonly believed that the newly proposed R_f metric addresses most of the inaccuracies of the old CRI and provides a predominantly satisfying solution as a scientifically accurate measure of color fidelity also in the context of white LED light sources. However, it still does not satisfy the need for a perception-related color quality measure that goes beyond fidelity. Despite being not intended for such a purpose, the CRI has been used for many years by lighting manufacturers and end users as a benchmark for designing and developing lighting solutions to achieve the critical balance between color quality (even though just partly reflected by the CRI) and energy efficiency. Still representing a pure color fidelity metric, the R_f metric, as stated by the CIE [152], is therefore not intended to replace the CRI in its current but often inappropriate application which would most likely not be better fulfilled by a more accurate fidelity index. Hence, there is obviously a strong need for a better extension or replacement of the CRI which on the one hand includes perception- and/or preference-related aspects of color quality but on the other hand still provides a practical solution for the industry.

3.2 MEMORY- AND PREFERENCE-BASED COLOR QUALITY METRICS

Despite the relatively large number of various proposals intended to replace the deficient CRI and to provide a future standard for describing the more perception-related aspects of color rendering, there are only four color quality metrics that are either based on memory or preferred colors and, therefore, specifically target the evaluation of light sources in terms of color preference and visual attractiveness. These are in chronological order Sanders' preferred color index R_p, Judd's flattery index $R_{flatt.}$, Thornton's color preference index R_{CPI}, and Smet's memory color rendition index R_m which should all be discussed in more detail in the following sections.

3.2.1 *Sanders' Preferred Color Index*

In 1959, Sanders published a series of psychophysical experiments intended to investigate the preferred colors and color tolerance ellipses of familiar real objects [25]. For these experiments, he made use of a cleverly designed viewing cabinet illustrated in Fig. 3.3 which enabled to vary the test objects' color appearance over a wide range while the visual adaptation state of the observers remained approximately constant.

As can be seen, the test object was placed on a wire mesh in the center of the viewing cabinet and could be observed through a small quadratic aperture in the front panel of the construction. Thus, a more or less fixed viewing angle was provided. The illumination of the test object was realized by three pairs of colored fluorescent lamps (green, blue, and pink) which were hidden in the lid of the viewing cabinet behind a diffusing screen. A specifically designed dimmer circuit was used to control the light emission of the three different channels so that light of variable color could be produced which homogeneously illuminated the test

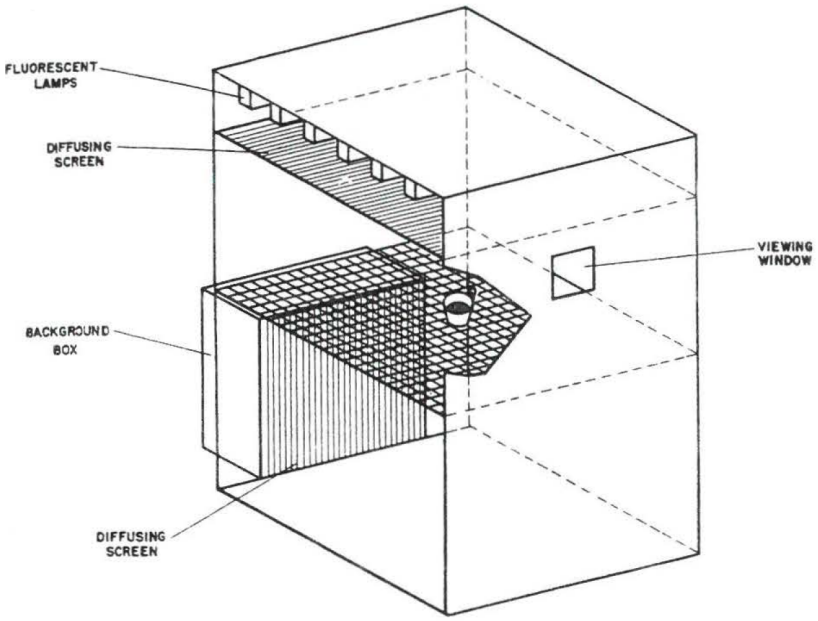


Figure 3.3 – Illustration of the experimental setup used by Sanders. Figure was taken from Ref. [25].

object under inspection and, consequently, caused a change in its color appearance. In order to maintain a constant adaptation state of the observers even when the object illumination was varied, a self-luminous white background which could be set to the chromaticities of either CIE illuminant B or CIE illuminant C was realized by using a setup of two additional fluorescent lamps installed behind a diffusing screen which partly replaced the lower part of the inner back panel of the viewing cabinet. Due to this self-luminous background in conjunction with the fixed viewing angle, observers basically had the impression that the test objects themselves changed color and that this effect is not caused by an external source. Hence, all clues regarding the actual object illumination could successfully be masked and, therefore, the observers' state of adaptation could be kept constant during the experiments.

In total, six different familiar objects were tested for both adaptation conditions. These objects – although very arguable in their relevance – were hand, face, tea, butter, beefsteak, and potato chips. In each case, at least 20 different light spectra were optimized for illuminating the respective test object to change its color appearance accordingly. The illuminance level at the test object's position was always constrained to be 215 lx. Based on their subjectively perceived preference, observers were asked to rate for each provided illumination condition the color appearance of the currently presented test object on a five-level rating scale running from "good" to "unsatisfactory". For calculation purposes, numbers from 100 to 0 in steps of 25 were additionally assigned to the individual semantic descriptors.

The results of the color appearance ratings for each combination of test object and adaptation condition were subsequently reported in terms of preferred color centers and corresponding chromatic tolerance ellipses. Figs. 3.4(a) and 3.4(b) summarize these results which were eventually used by Sanders in a further publication [26] to propose a preference-based color quality metric intended to evaluate the color rendering properties of white light sources without the need of defining an additional external reference.

Sanders' preferred color index R_p basically compares the CIE 1931 chromaticities of the six familiar test objects as perceived under the test illuminant and corrected by a translational chromatic adaptation transform of the Judd-type [174] with their experimentally determined

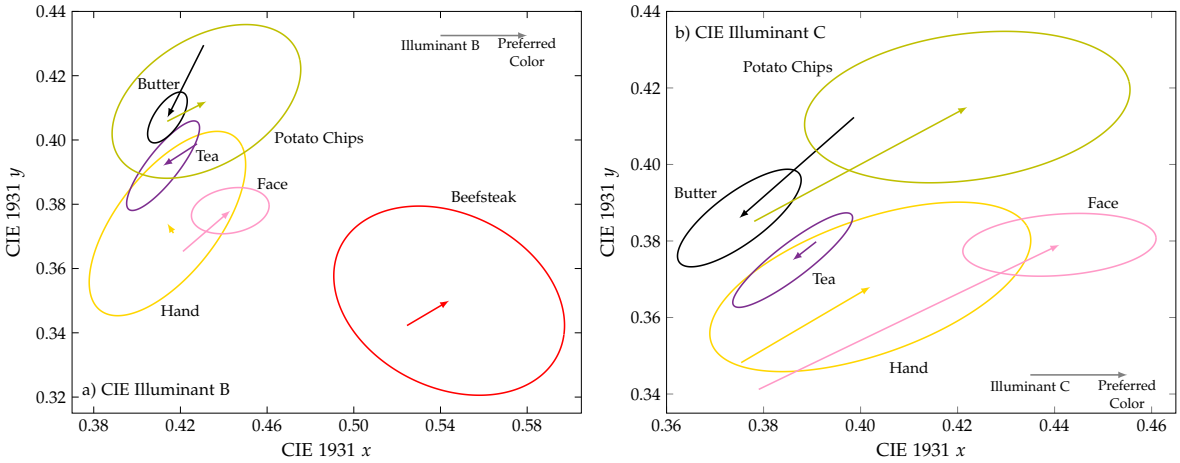


Figure 3.4 – Chromatic tolerance ellipses of the natural test objects assessed in Sanders' color appearance rating experiments for the two different adaptation conditions of CIE illuminant B (left figure) and CIE illuminant C (right figure). The tolerance boundaries were defined in such a way that they enclose all judgments with an average value larger than 50 and, therefore, represent a fair or better color rendition. Colored arrows shall indicate the direction and amount of the preference-induced chromaticity shifts in comparison to the respective test object's color appearance under reference illumination. Figures are reproduced from Ref. [25].

preferred chromaticities. Please note that in contrast to the CRI calculation, no reference illuminant was necessary any more introducing for the first time a new type of colorimetric measure which evaluates the color quality of a test light source with respect to what people expect or prefer certain objects to look like when illuminated by that light source.

Without the artificial incorporation of some kind of more or less arbitrarily chosen external reference, such preference or memory-based color quality metrics can in principle be considered to be an excellent choice for most lighting applications. The reason is simple: In the absence of an external reference, which is usually the case in a non-laboratory environment, object colors and, therefore, the color quality of a light source are always rated against the observers' notion of how the perceived objects should look like. In this sense, preference or memory-based rendition metrics are by definition best suited to account for this non-negligible cognitive component in the color quality evaluation of white light sources.

Coming back to Sanders' proposal, the comparison between the rendition results of a test light source and the preferred color centers is mathematically performed by calculating the so-called subjective color deviation

$$\mathrm{SCD}_i = \frac{\Delta S_i}{r_{\theta,i}} = \frac{\sqrt{\left(x_{\mathrm{t},i} - x_{\mathrm{p},i}\right)^2 + \left(y_{\mathrm{t},i} - y_{\mathrm{p},i}\right)^2}}{r_{\theta,i}} \tag{3.15}$$

for each of the six familiar test objects that were assessed in Sanders' original experiment. This measure gives the ratio between the difference ΔS_i of the test object's apparent (after chromatic adaptation) and preferred chromaticities and the radius $r_{\theta,i}$ of the corresponding tolerance ellipse measured in the same direction as the vectorial orientation of ΔS_i, where the color coordinates $(x_{\mathrm{t},i}, y_{\mathrm{t},i})$ and $(x_{\mathrm{p},i}, y_{\mathrm{p},i})$ denote the apparent and preferred chromaticities of the i^{th} test object, respectively. Please note that in order to reduce the error of the chromatic adaptation in calculating the SCD_i values, preferred chromaticity data of either CIE illuminant B or CIE illuminant C shall be chosen depending on whose white point is closest to the white point of the test light source.

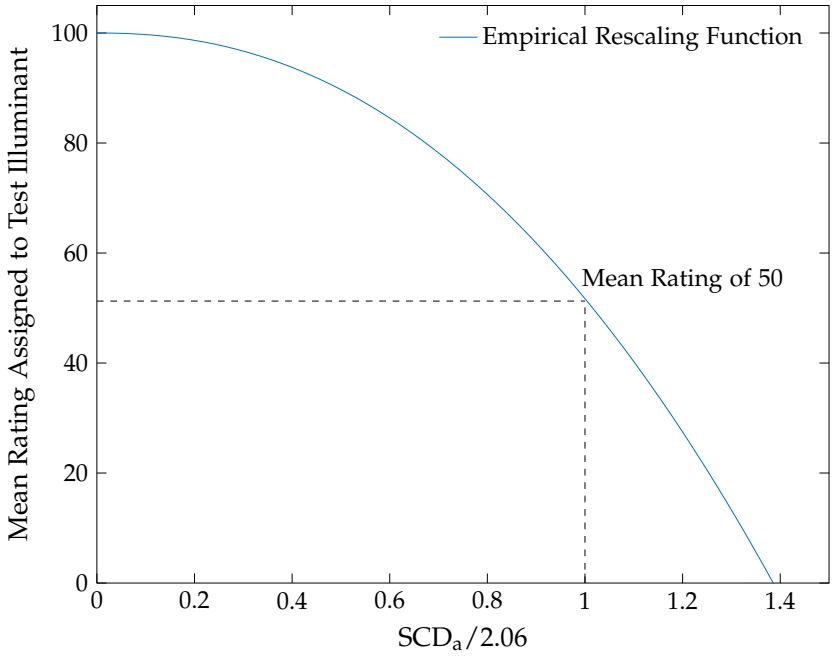


Figure 3.5 – Rescaling function to describe the relation between the calculated general SCD_a ratio and the mean preference-based color rendition rating assigned to the test light source. This functional relationship was empirically determined by Sanders based on his color appearance rating experiments. Figure is reproduced from Ref. [26].

According to Sanders, the above equation allows for a descriptive geometric interpretation. From this point of view, the SCD_i ratio indicates how close a certain test object's color appearance under the test illuminant given by its apparent chromaticities is located to the geometric center of the corresponding tolerance ellipse defining the preferred chromaticities. The smaller the ratio, the better the spatial agreement. Furthermore, in case that ΔS_i equals $r_{\theta,i}$ the respective ratio is 1, which means that the test object's apparent chromaticties reside exactly on the contour line of the corresponding tolerance ellipse representing by definition an average observer rating of 50 indicating a fair color rendition with respect to the observers' color preference.

Based on such a geometric interpretation of the SCD_i ratio, it can be shown that the following relation holds [83]:

$$(\mathrm{SCD}_i)^2 = g_{11,i}\left(x_{\mathrm{t},i} - x_{\mathrm{p},i}\right)^2 + g_{22,i}\left(y_{\mathrm{t},i} - y_{\mathrm{p},i}\right)^2 + 2g_{12,i}\left(x_{\mathrm{t},i} - x_{\mathrm{p},i}\right)\left(y_{\mathrm{t},i} - y_{\mathrm{p},i}\right), \qquad (3.16)$$

where the parameters $g_{kl,i}$ determine the size, shape, and orientation of the experimentally determined tolerance ellipses. The parameters for the different test objects assessed under both adaptation conditions are tabulated in Ref. [26] and, consequently, allow for an easy calculation of the specific SCD_i values for each of the six familiar test objects. A more general subjective color deviation SCD_a can eventually be obtained by averaging the individual SCD_i ratios.

However, as stated by Sanders [26], the unsuitability of an object color is not necessarily proportional to this SCD_a measure. Hence, an empirical relation between the general SCD_a ratio and the color appearance ratings observed in his experiments was derived, which is illustrated in Fig. 3.5. The result of this empirical rescaling defines Sanders' final R_p index. However, in order to assign a value of 85 to CIE illuminant B, as arbitrarily requested by Sanders, the SCD_a ratio has to be divided by a factor of 2.06 prior to rescaling. More details and further discussions can be found in Sanders' original publications [25, 26].

Table 3.1 – Summary of the 10 test samples used for the calculation of Judd's $R_{\text{flatt.}}$ indexed by their Munsell notation. The corresponding actual and preferred chromaticity coordinates for reference illuminant D65 in 1960 CIE-UCS chromaticity space are tabulated. Note that the weighting factors quoted by Judd are chosen more or less arbitrarily.

		ACTUAL		PREFERRED		
TEST SAMPLE #	MUNSELL NOTATION	u_a	v_a	u_p	v_p	WEIGHTS IN %
1	7.5 R 6/4	0.2351	0.3221	0.2451	0.3261	5
2	5 Y 6/4	0.2135	0.3429	0.2135	0.3429	15
3	5 GY 6/8	0.1817	0.3635	0.1717	0.3675	5
4	2.5 G 6/6	0.1570	0.3355	0.1470	0.3405	5
5	10 BG 6/4	0.1633	0.3045	0.1533	0.3025	5
6	5 PB 6/8	0.1732	0.2708	0.1672	0.2608	5
7	2.5 P 6/8	0.2076	0.2749	0.2116	0.2649	5
8	10 P 6/8	0.2378	0.2899	0.2478	0.2849	5
13	5 YR 8/4	0.2249	0.3292	0.2449	0.3402	35
14	5 GY 4/4	0.1862	0.3489	0.1662	0.3489	15

3.2.2 *Judd's Flattery Index*

According to Judd's original paper published in 1967 [27], his definition of a flattery index $R_{\text{flatt.}}$ for ligthing applications is intended to give *"a tentative measure of the degree to which the lighting installation produces the preferred colors of objects"*. Mathematically, it closely follows the calculation scheme of the CIE color rendering index as proposed in 1965 [16, 17] using 10 of the 14 Munsell test samples that were originally selected for testing color rendition.

However, instead of simply comparing the chromaticities of the test samples illuminated by a reference illuminant with those obtained for the test illuminant, Judd's flattery index takes into account the preferred chromaticities of the test samples, which were derived from the memory and preferred color data obtained by Newhall *et al.* [34], Sanders [25], and Bartleson [35], and weights them – more or less arbitrarily – according to their importance in lighting applications. In order to give an overview, Table 3.1 names the 10 Munsell test samples used for the calculation of $R_{\text{flatt.}}$ and summarizes their actual (u_a, v_a) and preferred (u_p, v_p) chromaticitiy coordinates for reference illuminant D65 in 1960 CIE-UCS chromaticity space. In addition, the corresponding weighting factors (in per cent) as quoted by Judd are tabulated. For a better visualization, Fig. 3.6 further illustrates the differences of the actual and preferred chromaticities. As can be seen, the preferred chromaticity shifts indicated by the red arrows generally point in the direction of increased chroma and, therefore, of higher saturation.

In order to allow for optimized light sources to perform better than the corresponding reference illuminants, Judd proposed to define the flattery index $R_{\text{flatt.}}$ in such a way that the reference illuminant is assigned a value of 90, whereas a light source that renders all of the n test samples to precisely match their preferred chromaticities is assigned a value of 100. Hence, Judd's flattery index may be written as

$$R_{\text{flatt.}} = 100 - 0.839\,\overline{\Delta E_{f,k}}, \tag{3.17}$$

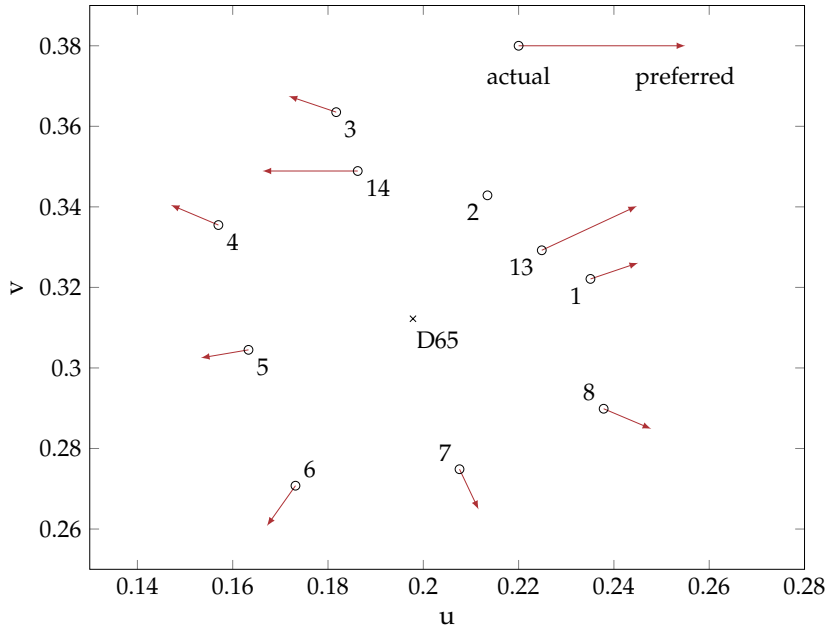


Figure 3.6 – Judd's preferred chromaticity shifts of the Munsell samples # 1-8, 13, and 14 for reference illuminant D65 shown on the 1960 CIE-UCS diagram. The chromaticity shifts are indicated by red arrows and generally point in the direction of increased chroma/saturation. Figure is reproduced from Ref. [27].

where $\overline{\Delta E_{f,k}}$ is the arithmetic mean of the weighted differences $\Delta E_{f,i}$ between the chromaticities of the $n = 10$ test samples illuminated by the test source and their preferred chromaticities under reference illumination. Here, $\Delta E_{f,i}$ is given by

$$\begin{aligned}\Delta E_{f,i} = 800 \Big\{ &\big[(u_{r,i} + \Delta u_{f,i} - u_r) - (u_{k,i} - u_k)\big]^2 \\ &+ \big[(v_{r,i} + \Delta v_{f,i} - v_r) - (v_{k,i} - v_k)\big]^2 \Big\}^{1/2},\end{aligned} \tag{3.18}$$

with $\Delta u_{f,i}$ and $\Delta v_{f,i}$ being the chromaticity shifts that have to be added to the chromaticity coordinates $(u_{r,i}, v_{r,i})$ of each test sample i illuminated by the reference light source having its white point at (u_r, v_r) in order to obtain the preferred chromaticities of the test samples. Furthermore, the variables $u_{k,i}$ and $v_{k,i}$ indicate the chromaticity coordinates of each test sample i illuminated by the test light source with (u_k, v_k) giving its respective white point.

As can be seen from Eq. (3.18), Judd's flattery index also incorporates a translational (or Judd-type) white-point adaptation transform to account for the chromaticity differences between the white points of the test and the reference illuminant [54]. By subtracting the white point chromaticities of the light sources from the corresponding chromaticity coordinates of the test samples only relative coordinates are considered. Note that although this is consistent with the calculation scheme of the CRI as defined in the 1960s, it significantly differs from the current definition of the CRI standardized in 1974 [158], which uses a more advanced von Kries-type transform to include white-point adaptation.

However, Eqs. (3.17) and (3.18) give only a preliminary definition of the flattery index. Due to criticism of Jerome and Nickerson [27], who stated that for providing a direct comparison between the CRI of a light source and its flattery index both measures should operate on the same scale, Judd decided to modify Eq. (3.17) in such a way that it exactly matches the CRI definition. Hence, the final version of Judd's flattery index reads

$$R_{\text{flatt.}} = 100 - 4.6\,\overline{\Delta E_{f,k}}, \tag{3.19}$$

i.e., the scaling factor was approximately increased by a factor of five. At the same time, this means that the chromaticity shifts of Eq. (3.18), which are illustrated in Fig. 3.6, must be reduced by the same factor in order to retain a flattery value of 90 to be still assigned to the reference illuminant. For further details and an overview of the reduced chromaticity shifts for each of the ten test samples upon which the $R_{\text{flatt.}}$ calculation is based, the interested reader is referred to Judd's 1967 paper [27].

3.2.3 *Thornton's Color Preference Index*

From a conceptional point of view, Thornton's color preference index (CPI) proposed in 1974 [30] is very similar to Judd's previously discussed flattery index. Instead of using actual preferred or memory colors like for example Sanders [26] and Smet *et al.* [23] did with their definition of a color quality metric, both Thornton's CPI and Judd's flattery index simply correct the chromaticities of a certain number of Munsell color samples illuminated by a reference light source with some experimentally motivated [25, 34, 35], but arbitrarily chosen preferred chromaticity shifts.

In contrast to Judd's flattery index, Thornton's CPI only uses the first eight Munsell samples given in Table 3.1 and keeps the original magnitude of the preferred chromaticity shifts illustrated in Fig. 3.6, i.e., no reduction by a factor of five is applied. In addition, all Munsell color samples are weighted equally so that the CPI as proposed by Thornton reads

$$R_{\text{CPI}} = 156 - 7.18\overline{\Delta E_{\text{k}}}, \tag{3.20}$$

where $\overline{\Delta E_{\text{k}}}$ is the arithmetic mean of the equally weighted differences $\Delta E_{\text{k},i}$ between the chromaticities of the eight test samples illuminated by the test light source and their preferred chromaticities under reference illumination. Equivalent to Judd's flattery index, these differences $\Delta E_{\text{k},i}$ are calculated by using Eq. (3.18). Furthermore, the scaling constants $a_{\text{CPI}} = 156$ and $b_{\text{CPI}} = 7.18$ in Eq. (3.20) were chosen in such way that i) a CPI value of 100 is assigned to the reference illuminant in order to be in accordance with the CRI definition and ii) the ratio of the constant b_{CPI} to the maximum CPI value is the same as the ratio of the CRI scaling constant $b_{\text{CRI}} = 4.6$ to its maximum value $R_{\text{a,max}} = 100$.

Please note that even though Judd's flattery index and Thornton's CPI including the corresponding chromaticity shifts are specifically designed for *"illuminants of source color not too different from that of D65"* [30], both color quality metrics will be applied over a wide range of light sources with different CCTs in a later part of this thesis. Hence, it is believed that replacing their outdated translational chromatic adaptation transform by for example the more sophisticated CAT02 approach would significantly enhance their predictive performance. For further details see Sec. 5.2.3.

3.2.4 *Smet's Memory Color Rendition Index*

Finally, Smet's memory color rendition index (MCRI) as the latest in a row of preference or memory-based color quality metrics should be discussed. In principle, the MCRI assesses the color quality of a light source by comparing the rendered color appearance of ten familiar test objects (green apple, ripe banana, orange, dried lavender, smurf®, strawberry yogurt, sliced cucumber, cauliflower, Caucasian skin, and a neutral gray sphere) with their actual memory colors, i.e., the color appearance observers expect the objects to look like in reality. These memory colors are defined by bivariate Gaussian similarity functions derived from experiment which describe the psychophysical response of observers to chromaticity devia-

tions from the corresponding memory color centers given by the centroids of these similarity distributions.

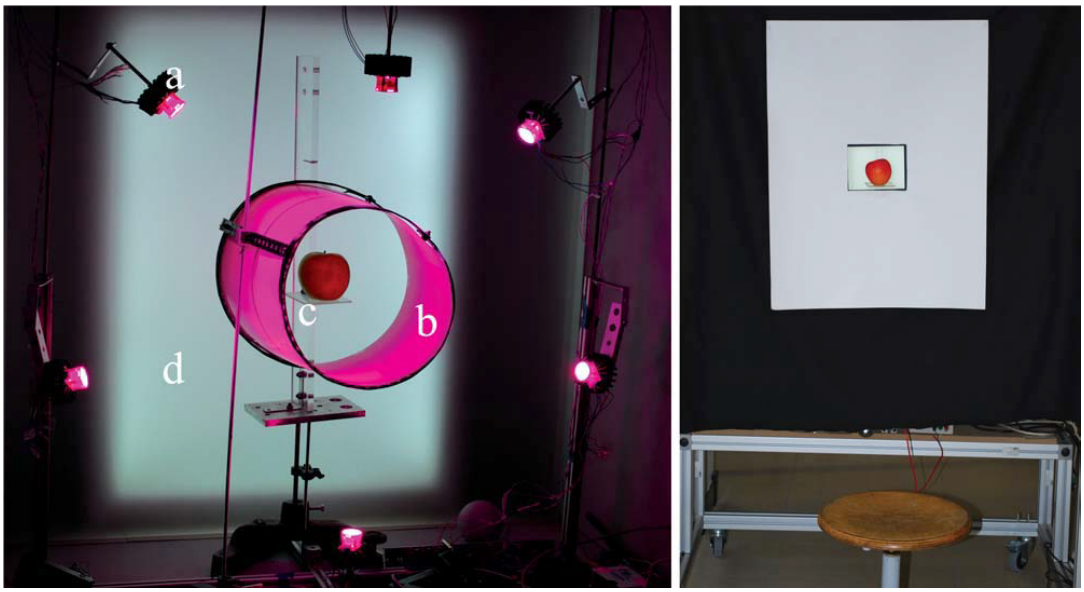


Figure 3.7 – Experimental setup used by Smet *et al.* [40]. Left image: Interior view of the illumination box showing a) PWM-controlled RGBA LED packages for illuminating the test object, b) diffusing tunnel for preventing specular reflections, c) transparent mounting for holding the test object in position, and d) self-luminous back panel for maintaining a constant adaptation state. Right image: Viewing cabinet as seen by the observer. The white front panel masks the inner life of the viewing cabinet and is illuminated by an external fluorescent lamp causing approximately the same chromaticity as exhibited by the back panel. Figures were taken from Ref. [40].

For the determination of the required Gaussian functions, a psychophysical color appearance rating experiment was devised by Smet *et al.* [40] in which various familiar test objects were presented to a group of observers in such a way that their color appearance could be varied over a wide range of possibilities while keeping the observers' chromatic adaptation state constant during the rating process. To achieve this goal, Smet *et al.*, similar to the approach of Sanders, made use of a specially designed illumination/viewing cabinet shown in Fig. 3.7 which masked all clues regarding the illumination that caused the change in the color appearance of the test objects. In order to maintain a constant adaption state, a self-luminous white back panel with a CCT of 5600 K, CIE 1964 chromaticity coordinates of $x_{10} = 0.3289$ and $y_{10} = 0.3514$, and a luminance of $590\,\mathrm{cd\,m^{-2}}$ was mounted on the inside of the viewing cabinet which was closed by a white front panel masking the inner life and leaving only a central 10° viewing aperture. By illuminating the front panel with an external fluorescent lamp, it could be guaranteed that the chromaticity coordinates of the front and the back panel were approximately the same. The uniformity of the object illumination, assessed by measuring the luminance at several locations across a neutral grey sphere which was placed inside the viewing cabinet, was reported to be $L_{\mathrm{min}}/L_{\mathrm{max}} = 0.90$.

For each familiar test object, more than 100 different individually optimized chromaticities with approximately equal luminances were rendered by adjusting the light output of the six radially arranged PWM-controlled RGBA LED packages which illuminated the inner part (where the test object is placed) of the viewing cabinet. In total, 32 observers took part in the experiments. They were asked to rate the presented test object's apparent color according to their idea of how the respective object should look like in reality adopting a five-point rating scale. Following the proposal of Yendrikhovskij *et al.* [38], bivariate Gaussian distributions

intended to represent an average observer were used to model for each test object the pooled observer ratings. An example for the test object of green apple is shown in Fig. 3.8(a).

a) Gaussian model fit to the mean observer ratings for the test object of green apple

b) Elliptic contours of the ten similarity functions representing a unit Mahalanobis distance

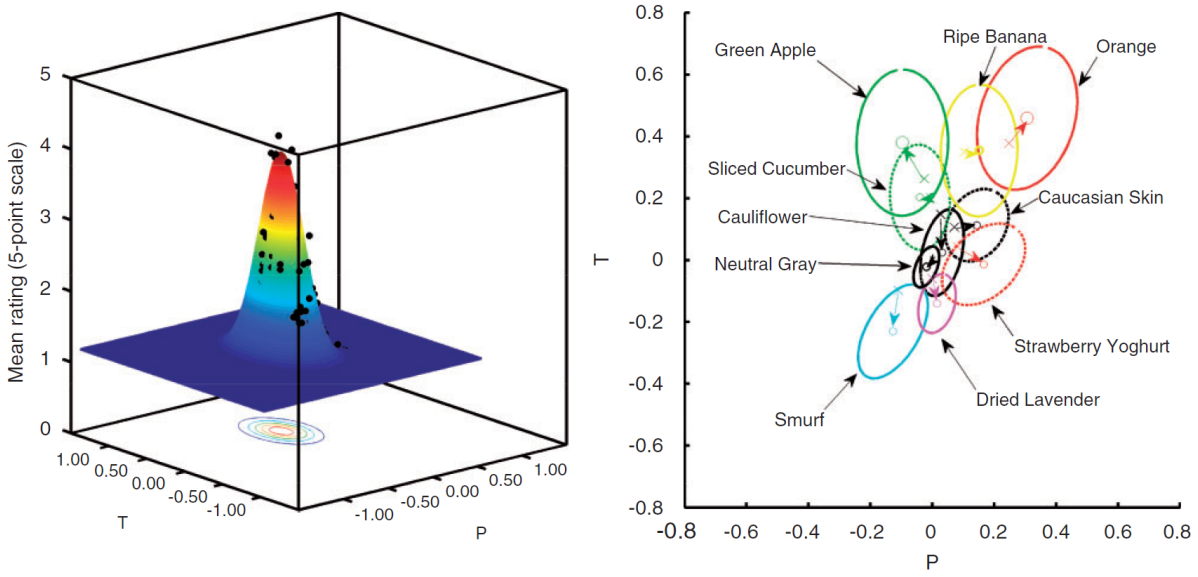


Figure 3.8 – Overview of the experimental results obtained by Smet *et al.* [40]. Left figure: Exemplary illustration of the bivariate Gaussian model fitted to the mean observer ratings for the test object of green apple. Right figure: Illustration of the elliptical cross-sections of the similarity functions of the ten different familiar test objects assessed in the experiments. Here, the contour lines represent a unit Mahalanobis distance [175, 176]. The memory color centers (centroids of the respective similarity functions) are given by the colored open circles, while the colored crosses mark the test objects' chromaticities as perceived under D65. Figures were taken from Ref. [54].

In addition Fig. 3.8(b) gives an overview of the elliptical cross-sections of the similarity functions of the ten different familiar test objects assessed in Smet *at al.*'s experiments. Each similarity function is simply obtained by normalizing the corresponding Gaussian fit and, therefore, can be described by a set of five different fitting parameters $a_{1,i}$ to $a_{5,i}$ defining its centroid, size, shape, and orientation. Hence, the similarity function of the i^{th} test object in IPT color space [177, 178] (i.e., the color space adopted by Smet *et al.*) is given by

$$S_i(P_i, T_i) = \exp\left(-\frac{1}{2}\left(a_{3,i}\left(P_i - a_{1,i}\right)^2 + 2a_{5,i}\left(P_i - a_{1,i}\right)\left(T_i - a_{2,i}\right) + a_{4,i}\left(T_i - a_{2,i}\right)^2\right)\right), \quad (3.21)$$

with the required parameters being tabulated in Ref. [53].

Based on these similarity distributions, Smet *et al.* [23, 52–54] derived their MCRI definition which is given as follows: For each of the ten familiar test objects illuminated by the test light source, the corresponding colors under D65 reference illuminant are calculated in IPT color space assuming a 10° standard observer (the spectral reflectances of the test objects needed for the calculation are also tabulated in Ref. [53]). Here, the CAT02 transform as incorporated in the CIECAM02 color appearance model is used to mathematically perform the chromatic adaptation. The required parameter D giving the degree of adaptation is either determined by the luminance of the adaptation field (see Eq. (2.26)) or, if the luminance is unknown, is recommended to be set to 0.9.

Next, the calculated object chromaticities (P_i, T_i) are inserted into the similarity functions of Eq. (3.21) in order to compute the degree of similarity between the test objects' apparent color appearance and their respective memory colors. The closer the function value $S_i(P_i, T_i)$

is to unity, the better the agreement. From these object-specific values, the general degree of similarity S_a is obtained by taking their geometric mean

$$S_a = \sqrt[10]{\prod_{i=1}^{10} S_i}. \tag{3.22}$$

Finally, the S_a value is rescaled to the standard CRI range defined from 0 to 100 allowing for a better comparison between these two metrics. This rescaling is achieved by a sigmoid function giving the final form of Smet's MCRI which reads

$$R_m = 100 \cdot \left(\frac{2}{\exp\left(1 + p_1^* |\ln S_a|^{p_2^*}\right)} \right)^{p_3^*}, \tag{3.23}$$

where the scaling parameters $p_1^* = 21.7016$, $p_2^* = 4.2106$, and $p_3^* = 2.4154$ were chosen in such a way that i) the CIE illuminants F4 and D65 have an R_m value of 50 and 90, respectively, and ii) light sources with $S_a < 0.5$ are assigned an MCRI of zero.

3.3 OTHER COLOR QUALITY METRICS

In addition to the preference- and memory-based color quality metrics of Sanders [26], Judd [27], Thornton [30], and Smet [23, 52–54] discussed in the preceding sections, several other color quality and rendition metrics will be used in a later chapter for comparing the predictive performances of all these metrics with a new proposal that should be devised throughout this thesis. The selection of additional metrics includes the CRI R_a [159], the general color fidelity index R_f [152], the gamut area index (GAI) [144], the arithmetic mean of GAI and R_a [153, 154], the color quality scale (Q_a,Q_f,Q_p,Q_g) [19], the CRI2012 [150], the feeling of contrast index (FCI) [146], and the IES TM-30 R_g measure [151].

Please note that this list, although far away from being complete, is assumed to represent a balanced cross-section of the three basic categories of color rendition which, according to Houser *et al.*'s review paper [179], are formed by fidelity-, preference-, and gamut-based color quality metrics, respectively. For further details on the concepts and implementation of the individual metric definitions, the interested reader is referred to the cited references.

REFERENCES OF CHAPTER III

[16] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.1, 1965.

[17] D. Nickerson and C. W. Jerome, "Color Rendering of Light Sources: CIE Method of Specification and its Application," *Illuminating Engineering*, vol. 60, no. 4, p. 262, 1965.

[19] W. Davis and Y. Ohno, "Color quality scale," *Optical Engineering*, vol. 49, no. 3, p. 033 602, 2010.

[23] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Memory Colours and Colour Quality Evaluation of Conventional and Solid-State Lamps," *Optics Express*, vol. 18, no. 25, pp. 26 229–26 244, 2010.

[25] C. L. Sanders, "Color Preferences for Natural Objects," *Illuminating Engineering*, vol. 54, p. 452, 1959.

[26] C. L. Sanders, "Assessment of Color Rendition under an Illuminant Using Color Tolerances for Natural Objects," *Illuminating Engineering*, vol. 54, p. 640, 1959.

[27] D. B. Judd, "A Flattery Index for Artificial Illuminants," *Illuminating Engineering*, vol. 62, p. 593, 1967.

[30] W. A. Thornton, "A Validation of the Color-Preference Index," *Journal of the Illuminating Engineering Society*, vol. 4, no. 1, p. 48, 1974.

[34] S. M. Newhall, R. W. Burnham, and J. R. Clark, "Comparison of Successive with Simultaneous Color Matching," *Journal of the Optical Society of America*, vol. 47, no. 1, p. 43, 1957.

[35] C. J. Bartleson, "Memory Colors of Familiar Objects," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 73, 1960.

[38] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Representation of Memory Prototype for an Object Color," *Color Research & Application*, vol. 24, no. 6, pp. 393–410, 1999.

[40] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Colour Appearance Rating of Familiar Real Objects," *Color Research & Application*, vol. 36, no. 3, pp. 192–200, 2011.

[52] K. Smet, S. Jost-Boissard, W. R. Ryckaert, *et al.*, "Validation of a Colour Rendering Index Based on Memory Colours," in *Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2010, pp. 136–142.

[53] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "A Memory Colour Quality Metric for White Light Sources," *Energy and Buildings*, vol. 49, p. 216, 2012.

[54] K. Smet and P. Hanselaer, "Memory and Preferred Colours and the Colour Rendition of White Light Sources," *Lighting Research & Technology*, vol. 48, no. 4, p. 393, 2016.

[73] Commission Internationale de l'Éclairage, *Colorimetry*, CIE Technical Report 15:2004, 3rd ed. 2004.

[83] D. L. MacAdam, "Specification of Small Chromaticity Differences," *Journal of the Optical Society of America*, vol. 33, no. 1, pp. 18–26, 1943.

[144] J. P. Freyssinier and M. Rea, "A Two-metric Proposal to Specify the Color-rendering Properties of Light Sources for Retail Lighting," in *Proceedings of the SPIE 7784, 10th International Conference on Solid State Lighting*, San Diego, CA, USA: International Society for Optics and Photonics, 2010, pp. 77840V–1–77840V–6.

[145] M. Rea, L. Deng, and R. Wolsey, "Full-Spectrum Light Sources," *Lighting Answers*, vol. 7, no. 5, pp. 1–18, 2005.

[146] K. Hashimoto, T. Yano, M. Shimizu, *et al.*, "New Method of Specifying Color Rendering Properties of Light Sources Based on Feeling of Contrast," *Color Research & Application*, vol. 32, no. 5, pp. 361–371, 2007.

[147] W. A. Thornton, "Color-discrimination Index," *Journal of the Optical Society of America*, vol. 62, no. 2, pp. 191–194, 1972.

[148] S. A. Fotios, "The Perception of Light Sources of Different Colour Properties," PhD thesis, University of Manchester Institute of Science and Technology, 1997.

[149] M. R. Luo, "The Quality of Light Sources," *Coloration Technology*, vol. 127, no. 2, pp. 75–87, 2011.

[150] K. A. G. Smet, J. Schanda, and L. Whitehead, "CRI2012: A Proposal for Updating the CIE Colour Rendering Index," *Lighting Research & Technology*, vol. 45, no. 6, pp. 689–709, 2013.

[151] A. David, P. T. Fini, K. W. Houser, *et al.*, "Development of the IES Method for Evaluating the Color Rendition of Light Sources," *Optics Express*, vol. 23, no. 12, pp. 15 888–15 906, 2015.

[152] Commission Internationale de l'Éclairage, "Colour Fidelity Index for Accurate Scientific Use," *CIE Technical Report 224*, 2017.

[153] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Correlation Between Colour Quality Metric Predictions and Visual Appreciation of Light Sources," *Optics Express*, vol. 19, no. 9, pp. 8151–8166, 2011.

[154] S. Jost-Boissard, P. Avouac, and M. Fontoynont, "Assessing the Colour Quality of LED Sources: Naturalness, Attractiveness, Colourfulness and Colour Difference," *Lighting Research & Technology*, vol. 47, no. 7, pp. 769–794, 2015.

[155] P. Bodrogi, S. Brúckner, and T. Q. Khanh, "Ordinal Scale Based Description of Colour Rendering," *Color Research & Application*, vol. 36, no. 4, pp. 272–285, 2011.

[156] L. A. Whitehead and M. A. Mossman, "A Monte Carlo Method for Assessing Color Rendering Quality With Possible Application to Color Rendering Standards," *Color Research & Application*, vol. 37, no. 1, pp. 13–22, 2012.

[157] P. van der Burgt and J. van Kemenade, "About Color Rendition of Light Sources: The Balance Between Simplicity and Accuracy," *Color Research & Application*, vol. 35, no. 2, pp. 85–93, 2010.

[158] C. I. de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.2, 1974.

[159] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.3, 1995.

[160] C. Li, M. R. Luo, and C. Li, *Assessing Colour Rendering Properties of Daylight Sources Part II: A New Colour Rendering Index: CRI-CAM02UCS*, https://www.researchgate.net/publication/268364092_Assessing_Colour_Rendering_Properties_of_Daylight_Sources_Part_II_A_New_Colour_Rendering_Index_CRI-CAM02UCS, [Online; accessed 16-April-2018], 2009.

[161] H. Yaguchi, Y. Takahashi, and Y. Mizokami, "Categorical Colour Rendering Index Based on the CIECAM02," in *Proceedings of the 12th AIC Colour Congress*, Newcastle Gateshead, United Kingdom: Association Internationale de la Couleur (AIC), 2013, pp. 1441–1444.

[162] G. Wyszecki, "Proposal for a New Color-Difference Formula," *Journal of the Optical Society of America*, vol. 53, no. 11, pp. 1318–1319, 1963.

[163] Commission Internationale de l'Éclairage, *CIE Position Statement on CRI and Colour Quality Metrics*, http://www.cie.co.at/sites/default/files/CIE%20Position%20Statement%20on%20CRI%20and%20Colour%20Quality%20Metrics%20V2.pdf, [Online; accessed 16-April-2018], 2015.

[164] Commission Internationale de l'Éclairage, "Colour Rendering of White LED Light Sources," *CIE Technical Report 177 : 2007*, 2007.

[165] P. Bodrogi, P. Csuti, F. Szabó, *et al.*, "Why Does the CIE Colour Rendering Index Fail for White RGB LED Light Sources?" In *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 24–27.

[166] Y. Ohno, "Simulation Analysis of White LED Spectra and Color Rendering," in *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 28–32.

[167] Y. Ohno, "Spectral Design Considerations for Color Rendering of White LED Light Sources.," *Optical Engineering*, vol. 44, no. 11, pp. 111 302–111 310, 2005.

[168] W. Davis, J. L. Gardner, and Y. Ohno, "NIST Facility for Color Rendering Simulation," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 519–522.

[169] Y. Nakano, H. Tahara, K. Suehara, *et al.*, "Application of Multispectral Camera to Color Rendering Simulator," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 1625–1628.

[170] N. Sándor and J. Schanda, "CIE Visual Colour-rendering Experiments," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 511–514.

[171] N. Sándor, P. Csuti, P. Bodrogi, *et al.*, "Visual Observation of Colour Rendering," in *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 16–19.

[172] N. Sándor, P. Bodrogi, P. Csuti, *et al.*, "Direct Visual Assessment of Colour Rendering," in *Proceedings of the 25th CIE Session*, Vienna, Austria: International Commission on Illumination CIE, 2003, pp. D1–42–D1–45.

[173] F. Szabó, N. Sándor, P. Bodrogi, *et al.*, "Colour rendering of white LED light sources: Visual experiment with colour samples simulated on a colour monitor," in *Proceedings of the CIE* 2005 *Midterm Meeting: Vision and Lighting in Mesopic Conditions*, Vienna, Austria: International Commission on Illumination CIE, 2005.

[174] D. B. Judd, "Hue Saturation and Lightness of Surface Colors with Chromatic Illumination," *Journal of the Optical Society of America*, vol. 30, no. 1, pp. 2–32, 1940.

[175] P. C. Mahalanobis, "On the Generalised Distance in Statistics," *Proceedings of the National Institute of Science in India*, vol. 2, no. 1, pp. 49–55, 1936.

[176] M. Sapp, F. Obiakor, A. J. Gregas, *et al.*, "Mahalanobis Distance: A Multivariate Measure of Effect in Hypnosis Research," *Sleep and Hypnosis*, vol. 9, no. 2, pp. 67–70, 2007.

[177] F. Ebner, "Derivation and Modelling Hue Uniformity and Development of the IPT Color Space," PhD thesis, Rochester Institute of Technology, 1998.

[178] F. Ebner and M. D. Fairchild, "Development and Testing of a Color Space (IPT) with Improved Hue Uniformity," in *Proceedings of the 6th Color and Imaging Conference: Color Science, Systems and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 1998, pp. 8–13.

[179] K. W. Houser, M. Wei, A. David, *et al.*, "Review of Measures for Light-source Color Rendition and Considerations for a Two-measure System for Characterizing Color Rendition," *Optics Express*, vol. 21, no. 8, pp. 10 393–10 411, 2013.

[180] J. A. C. Yule, *Principles of Color Reproduction: Applied to Photomechanical Reproduction, Color Photography, and the Ink, Paper, and Other Related Industries*. New York, NY, USA: John Wiley & Sons Inc., 1967.

Part 4

COLOR APPEARANCE RATING OF FAMILIAR OBJECTS

"Color was an art long before it was a science, and consequently the language of color is poetic rather than factual"

– John A. C. Yule [180]

4

COLOR APPEARANCE RATING OF FAMILIAR OBJECTS

The concept of memory colors was first introduced by Hering [32] in the late 19th century in order to explain his findings of color constancy in visual perception. Following Hering's original definition, the term "memory color" is used to describe the typical color an individual observer has in mind when thinking of or looking at certain familiar objects that he or she acquired frequently and memorized stably through his or her experience with the respective objects. Hering further stated that the spectacles of memory, i.e., the prior knowledge of how the representative color should look like, might induce a considerable shift regarding the perceived color appearance of a familiar object.

Several subsequent studies investigating the influence of memory on the perception of object colors validated Hering's claim by showing that the expectation of the color of an object being impressed on the observer's long-term memory undeniably affects its perceived color [25, 31, 34, 35, 37, 38, 40, 42, 43, 181–185]. Independent experiments performed by Duncker [181], Bruner *et al.* [182], and Harper [183] for example revealed that the memory color, which is associated with a certain familiar object, has a significantly measurable impact on the observers' color matching results when the object is presented in such a way that its actual color is not the color characteristically identified with it, i.e., a yellow banana illuminated by a red light source leads to a more yellowish color matching result than its actual color of presentation would suggest.

In another series of display-based experiments, Hansen *et al.* [43] and Olkkonen *et al.* [185] provided further evidence of the memory induced modulation of human color perception. Applying the method of adjustment, they asked several subjects to interactively manipulate the rendered color appearance of natural fruit objects presented on the monitor as digitized photographs until they appeared achromatic. It was found that these familiar objects were perceived to be gray when their color was slightly shifted away from the actual neutral grey point into the opposite direction of their typical color. Hence, the color of for example a yellow banana was adjusted to an opponent slightly blueish hue in order to appear achromatic for the observers. These results indicate that natural and other familiar objects always tend to be perceived in their typical memorized color. By providing cues to the visual system which may serve as an internal reference, memory colors are therefore considered to play a crucial role for the mechanisms of color constancy and color appearance under varying illumination conditions [32, 43, 44, 55, 185] as well as for color reproduction [33, 39, 45–51] and color rendering applications [23, 26, 27, 30, 52–54].

Because of their potential use as an internal reference in the assessment of color appearance and color quality, first experiments aiming to explore the characteristics of memory colors quantitatively were performed by Newhall *et al.* [34] and Bartleson [35]. Both studies experimentally determined the memory colors of various familiar objects and their corresponding acceptance boundaries using a large number of Munsell color patches which were presented to the observers in a viewing booth. Given only the name of a certain familiar object (e.g., green grass, skin, red brick, blue sky,...), the participants were asked to choose from the set of Munsell color patches the one which best represents their idea of how the color of the corresponding object looked like in reality.

When comparing these selections with the average measured colors of the corresponding test objects [24, 36], significant differences in hue, chroma, and lightness could be observed. Memory lightness for example averaged higher by slightly more than one Munsell step, while memory hues tend to move towards the typical or dominant hues commonly associated with the test objects. Furthermore, most of the tested object colors showed increased chroma when recalled from memory, an overall trend which has been confirmed in various subsequent experiments [25, 31, 33, 37–41]. However, some of the test objects of naturally low saturation like sand, skin, and concrete, did not show this effect, which was also confirmed by the findings of Pérez-Carpinell *et al.* [37] who examined the memory induced shifts in dominant wavelength and colorimetric purity of eight familiar objects with the help of Munsell color patches assessed under different illuminants (D65 and A) while considering two different observer groups with varying degree of expertise regarding the technical and artistic aspects of color. Again, only the names of the test objects were given as a hint to the observers that were asked to perform the color matching with respect to their long-term memory impression of the respective object.

By excluding object properties such as texture, shape, and size – which were demonstrated to have an impact on both similarity ratings [38, 186] and color constancy [187, 188] – from the assessments, these memory color matching experiments based on abstract Munsell color patches could only lead to a rough estimate of the underlying ground truth rather than providing a detailed analysis of how people assess memory colors under realistic viewing conditions. In order to overcome these shortcomings, Sanders [25] in the late 1950s was the first who examined color appearance ratings with respect to the observers' long-term memory by presenting them real familiar test objects – an approach which was later adopted and refined in a seminal work by Smet *et al.* [40]. Details on the concept of both studies and the deduced results can be found in Secs. 3.2.1 and 3.2.4, respectively.

In both cases, the test objects were presented to the observers in a cleverly designed viewing cabinet in such a way that the object's color appearance could be varied over a wide range while the visual adaptation level of the observers remained approximately constant. For each test object, a certain variety of different illumination conditions could then be optimized leading to a perceivable change in its color appearance. During the experiments, the participants were asked to rate the similarity of the currently presented object color with their idea of what the object looked like in reality on a five-level scale of the Likert-type [189]. Besides providing color centers and acceptance boundaries for the memory colors of each of the test objects, such a rating scale also allows for the introduction of certain weighting functions describing the quantitative assessment of the color appearance of the test objects and is therefore considered to be a superb choice that should also be applied for the experiments performed within the scope of this thesis. The so-obtained weighting functions could eventually be used, similar to the approaches of Sanders [26] and Smet *et al.* [23, 52–54], to device a new algorithm for the improved memory-based evaluation of the color rendering properties of white light sources, which basically should be the main goal of the present work.

Even though the experiments of Sanders [25] and Smet *et al.* [40] provide a quite good methodology for quantifying the assessment of memory colors, they both lack the possibility of taking into account the influence of realistic adaptation conditions on the observer ratings. Hence, a new experiment on the assessment of memory colors providing more realistic viewing and adaptation conditions was conducted as part of this thesis, whose conceptual design and obtained results should be reported here. Besides giving an overview of the experimental setup, the test color selection, and the composition of the panel of subjects, a thorough statistical analysis of the observers' color appearance ratings will additionally be provided in the following sections.

4.1 DISPLAY-BASED METHODS OF MEMORY COLOR EVALUATION

Before considering in detail the setup and results of the newly devised experiments presented in the current thesis, a selection of display-based methods for the evaluation of memory colors reported by various researchers should be discussed first. In this context, special emphasis will be put on the drawbacks of such methods regarding their quite small benefit for an application to real lighting scenarios on which the focus of the present work should be. The following overview is limited to studies that assessed memory colors by performing color appearance rating or matching experiments of familiar objects similar to those reported later in this thesis, with the only difference that the former were conducted with the help of a computer-driven display visualization and not by presenting real objects to the observers.

YENDRIKHOVSKIJ *et al.* [38] (1999) In this study, the authors used a display-based setup to investigate the characteristics of how people rate the color appearance of a specific test object presented to them in a huge variety of different colors. Due to its general familiarity, a ripe (yellow) banana was chosen as the test object. This object was then presented to the observers in form of digitally processed images on a colorimetrically characterized and calibrated cathode ray tube (CRT) monitor adjusted to D55 reference white. Three different situations of image content shown in Fig. 4.1 were prepared in order to examine whether the

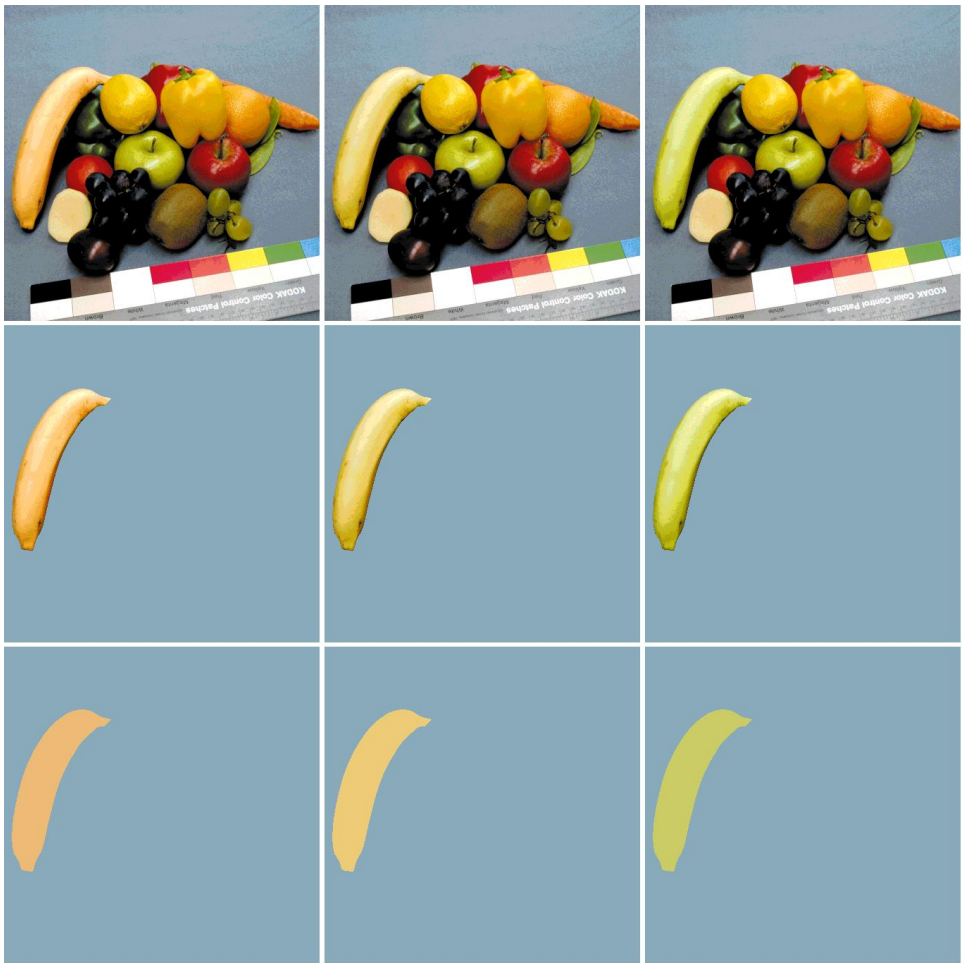


Figure 4.1 – Examples of the test object's images rated by the observers. Three different situations were examined which are the banana presented in a multi-fruit environment (upper row), the single-fruit case (middle row), and the banana presented as colored contours (lower row). Figures were taken from Ref. [38].

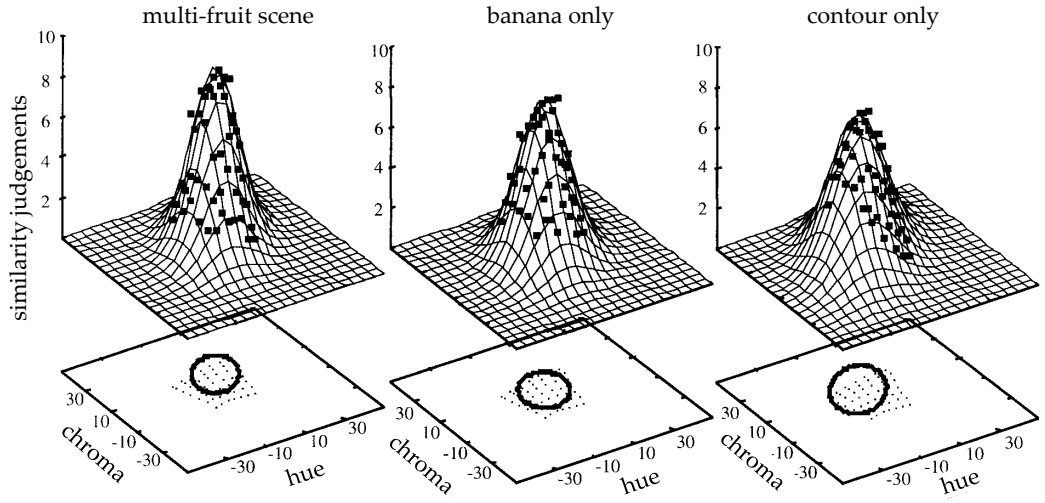


Figure 4.2 – Gaussian modeling of the observers' average similarity judgements (filled squares) regarding the color appearance of various banana samples presented on a CRT monitor in comparison to the observers' idea of how a ripe banana should typically look like in reality. Results for the three differnet types of image content are reported. Figure was taken from Ref. [38].

naturalness of the object depiction and the complexity of the image content have an impact on the color appearance ratings of the observers.

Hence, for each of the three situations, 49 color variations were created by pixel-wisely manipulating the color attributes of chroma and hue of the banana object without changing its lightness correlate or the rest of the image content. A total number of eight subjects took part in the experiments. For each presented test image, the participants were asked to judge the similarity in colors between the banana samples' color appearance displayed on the CRT monitor and a typical ripe banana as they remembered it from their memory. The rating was performed on an eleven-point numerical category scale ranging from 0 (no similarity) to 10 (complete agreement) and was repeated twice for all color variations of each situation of image content. The respective similarity judgments of the eight observers were averaged and subsequently visualized in a chromaticity space spanned by the average hue and chroma correlates of the banana samples which was extended by a third dimension representing the observers' mean rating values. In Fig. 4.2, the corresponding results, which could be modeled by bivariate Gaussian functions, are illustrated for the three different types of image content.

Further analysis of the Gaussian functions and of their perpendicular projection to the chromaticity plane revealed that the centroids and spread of the bivariate Gaussian distributions, representing the prototypical memory color of a ripe banana in the given chromaticity space, significantly varied between the different types of banana depiction, while their shape and orientation remained more or less the same. This observed variation appeared to happen in such a way, with the multi-fruit scenario showing the smallest spread in the corresponding similarity distribution, that the authors were tempted to conclude that *"the naturalness of the banana depiction facilitates the similarity judgments between the apparent and prototypical colors"*. Moreover, the memory color centers/locations defined by the centroids of the respective Gaussian similarity distributions were found to exhibit slightly more saturation than the reference representation of the banana samples would have suggested, which, as stated previously, can be considered as a general feature of memory color representations of familiar objects.

TARCZALI *et al.* [41] (2006) Based on the findings reported by Yendrikhovskij *et al.* [38], a new study was devised exploring the memory color assessments of six different familiar objects in a three-phase color matching experiment performed on a colorimetrically characterized CRT display. Two different cultural observer groups, consisting of Hungarian and Korean observers, respectively, were tested and compared with each other. In total, eleven Hungarian and nine Korean subjects took part in the experiments which were conducted in a completely dark room with the display's white point set to 6500 K. In the first phase of the experiments, a method of choice was applied where a single color name of one of the six familiar objects (Caucasian skin, blue sky, green grass, oriental skin, deciduous foliage, and orange) was given to the observers who had to choose from a 4×4 array of displayed color patches the one that best fitted their idea of how the respective object looked like in reality. A total number of thirty randomized repetitions of this memory color matching procedure were performed for each observer and test object and the resulting memory color centers and tolerance radii defined as twice the standard deviations of the mean memory colors were reported.

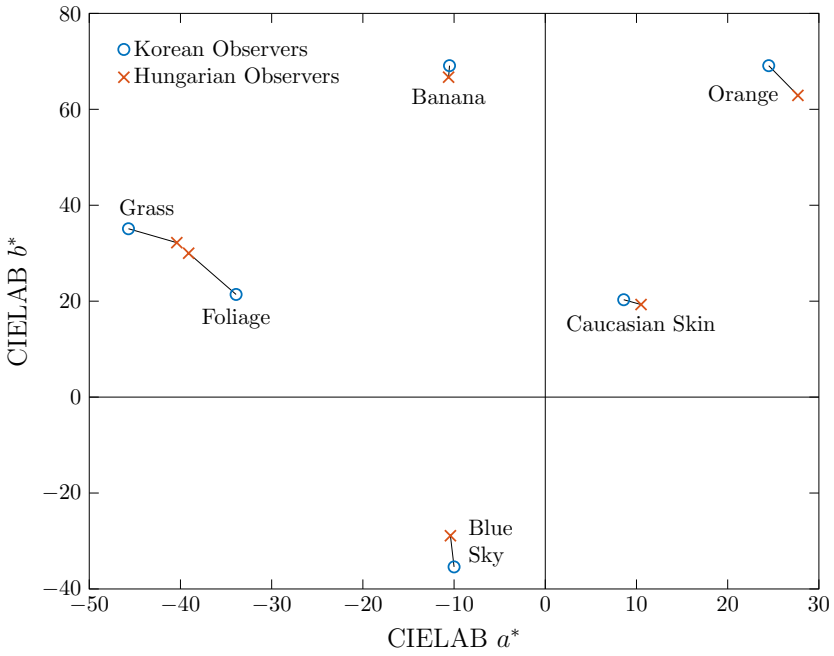


Figure 4.3 – Inter-cultural comparison of the long-term memory color representations of the six different test objects obtained from averaging the Korean and Hungarian observer assessments of all three experimental phases and all repetitions. Figure was reproduced from Table 3.9 of Ref. [190].

In the second phase of the experiments, the method of reproduction was applied. A square-shaped color patch of changeable color was presented to the observers on the CRT display with the color name of one of the familiar test objects being written above the color patch. The order in which the color names were displayed for the individual observer was again randomized. During the experiment, three sliders could be used to adjust the hue, saturation, and lightness of the color patch in such a way that it matched the observer's notion of how the color appearance of the respective test object should be in reality. Ten repetitions were performed for each test object and the mean findings for each observer were reported.

The third phase was similar to the second one, with the difference that this time no color name was given. Instead, the color patch was shown as part of a grey-scale photorealistic image displayed on the CRT monitor which could unambiguously assigned to one of the familiar test objects. The three sliders could again be used to adjust the color features of the

color patch to reproduce the most suitable color corresponding to that part of the grey-scale image where the changeable color patch was displayed. Again, the matching was repeated ten times for each test object in randomized order and the mean results for each observer were reported.

Based on the findings of the three phases, an overall mean long-term memory color was calculated for each test object and observer group. Hence, the mean results of both Hungarian and Korean observers could be compared allowing for drawing some conclusions about the impact of the cultural background on memory color assessments. In Fig. 4.3, the CIELAB chromaticities of the memory color centers are shown for both observer groups. As can be seen, slight deviations between Hungarian and Korean observers occur for all test objects indicated by solid line segments. When considering the distributions of the color coordinates L^*, a^*, and b^* of both cultural observer groups around their respective means separately, significant differences between Hungarian and Korean observers with respect to at least one of these coordinates could be reported, with the exception of blue sky, for all familiar test objects adopting a 5 % significance level. Even though this is a very arguable way of dealing with multivariate data, which actually does not allow for drawing further conclusions about cultural dependencies, Tarczali *et al.* [41, 190], based on these findings, reported a certain cultural impact on the memory color assessments. However, due to the poor way of performing the statistical analysis, such a conclusion must be considered as very doubtful.

SMET *et al.* [55] (2014) In this study, the effect of cross-cultural differences on the color appearance ratings and memory colors of familiar test objects, as being reported in a first attempt by Tarczali *et al.* [41], was further investigated by using a liquid-crystal-display(LCD)-based methodology specifically developed for being applied worldwide. In total, results from seven different research institutes of seven different countries could be collected. These countries are Belgium, Hungary, Brazil, Colombia, Taiwan, China, and Iran. Even though each laboratory used its own equipment for running the experiments, the established methodology provided an exact instruction how to calibrate and characterize the LCD devices as well as how to gather the observers' rating data so that the subsequent analysis could be performed on a common ground.

In each country, a set of eleven different familiar objects (green apple, ripe banana, ripe lemon, cauliflower, orange, strawberry, tomato, dried lavender, smurf®, Caucasian skin, and Asian skin), being well distributed around the hue circle, were presented as digitally processed images to a certain number of observers varying from country to country. The displays' white points were set to D65 chromaticity and a luminance of approximately 200 cd m^{-2}. Similar to the study of Yendrikhovskij *et al.* [38], over 100 different color variations were created for each test object by accordingly manipulating the chromatic pixel content, while keeping the lightness correlate unchanged. These processed images were subsequently displayed on the LCD and the observers were asked to rate the apparent color of the presented object according to what they thought it should look like in reality.

Again, the average rating results for each region could be modeled by using bivariate Gaussian similarity distributions. Although statistical testing revealed significant differences between the region average observers and a global observer obtained by pooling the data of a specific test object from all regions, the effect size of the cultural background was found to be very small. For illustrative purposes, Fig. 4.4 visualizes the cross-regional differences for the green apple which statistically showed the largest effect size of all test objects. As can be seen, slight deviations between the different countries/regions are present which, as stated by the authors, might be due the object's naturally occurring large variations in color.

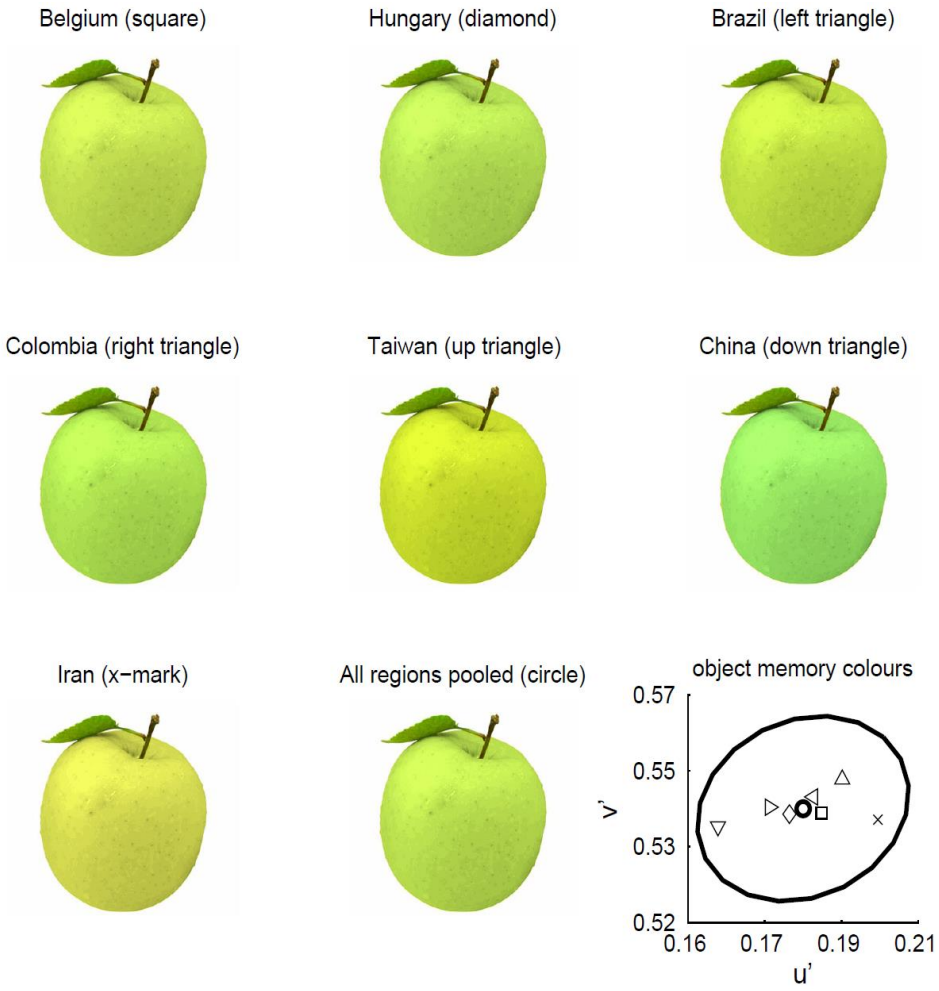


Figure 4.4 – Illustrations of the region average and the global memory colors for the test object of green apple. The (u', v') subplot in the last row of the right coloumn depicts the differences among the regional memory color centers. In addition, the memory color center of the global observer is shown together with the corresponding tolerance boundary of color acceptibility. Figures were taken from Ref. [55].

However, from the subplot of Fig. 4.4 it should be noted that all regional memory color centers still lie within the tolerance boundary ellipse assigned to the global observer which basically denotes the observer's range of color acceptability. Hence, bearing in mind that for the remaining test objects the effect size of the impact of the cultural background on the color appearance ratings was observed to be much smaller leading to differences between the regional and global memory color centers to be of the same order of magnitude or even smaller than the typical inter-observer variabilities, the authors concluded that even though statistical significance was obtained, the regional differences were likely to be of no practical importance, which partly contradicts the findings of Tarczali *et al.* [41, 190] discussed above. However, on the other hand, it must be expected that Tarczali *et al.* would have come to the same conclusions if they had performed a more sophisticated statistical analysis capable of treating multivariate data properly, like Smet *et al.* [55] did in their research.

ZHU *et al.* [191] (2017) In order to address the inconsistencies between the results of Tarczali *et al.* [41, 190] and those obtained by Smet *et al.* [55], a new display-based experiment was devised by an international Chinese-German research collaboration of which the author of this thesis was also part of. Please note that some of the results reported here were also

published elsewhere [192, 193]. In contrast to the study of Smet *et al.* [55], the exact same experimental equipment – a high-resolution AdobeRGB LED backlight display – was used in both countries for data acquisition. Special care was taken to develop a common methodology ensuring a proper display calibration and characterization yielding the same, colorimetrically accurate performance in both countries. As before, the display's white point was set to D65 chromaticity with a slightly smaller luminance of 110 cd m^{-2} when compared to the approach of Smet *et al.* [55].

The subsequent experiments basically comprised two different parts both investigating the impact of inter-cultural differences between Chinese and German observers on the assessment of memory colors using a set of 18 familiar test objects (banana, strawberry, orange, kiwi, nectarine, green apple, broccoli, lemon, aubergine, lavender, red rose, pink lotus, green grass, blue jeans, blue sky, Caucasian skin, oriental skin, and African skin) which were partly identical to those applied in the other studies. In the first part, the same method of reproduction was applied as reported by Tarczali *et al.* [41] for their second phase, i.e., a square-shaped color patch of changeable color was presented to the observers on the LCD with the color name of one of the familiar test objects being written above the color patch (see Fig. 4.5). The order in which the color names were displayed for the individual observer was randomized. In order to adjust the appearance of the color patch to match the observer's respective memory color, three different sliders were again provided to manipulate the patch's CIELAB correlates of lightness, chroma, and hue, accordingly. The average results of 25 Chinese and 44 German observers were reported in terms of regional memory color centers and corresponding matching tolerance ellipses.

In the second part of the experiments, more or less the same experimental setup and procedure was used with the only difference that this time a grey-scale image representation of

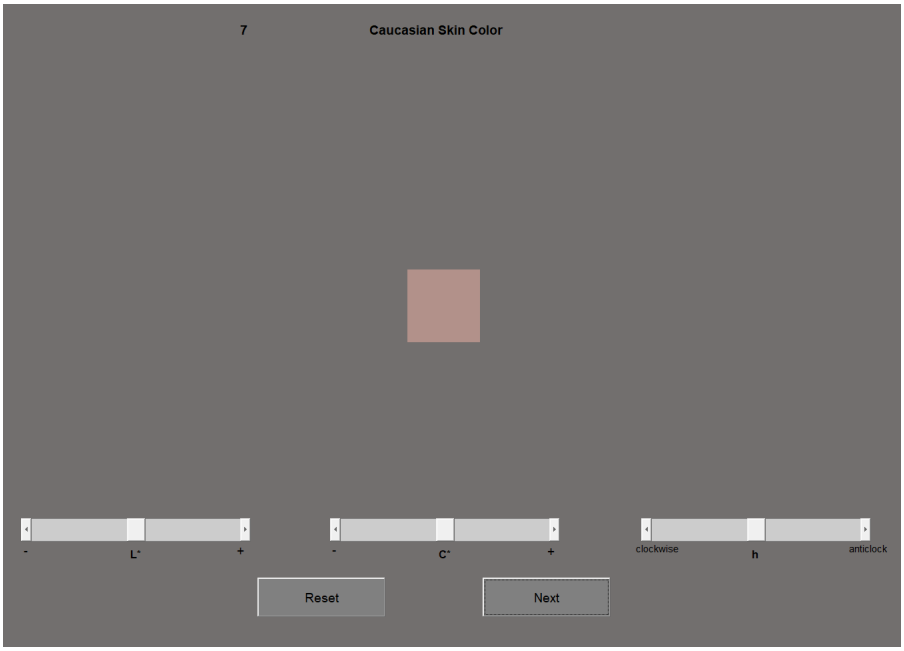


Figure 4.5 – Layout of the graphical user interface for manipulating the color appearance of the square-shaped homogeneous color patch. In the first part of the experiment only the test object's name, as illustrated here, was given as a clue to the observers who were asked to adjust the CIELAB lightness, chroma, and hue correlates of the color patch by moving the corresponding three sliders until its color appearance matched their idea of how the respective test object should look like in reality. In the second part of the experiment, a grey-scale image representation of the test object providing an additional clue was displayed between the color patch and the test object's name.

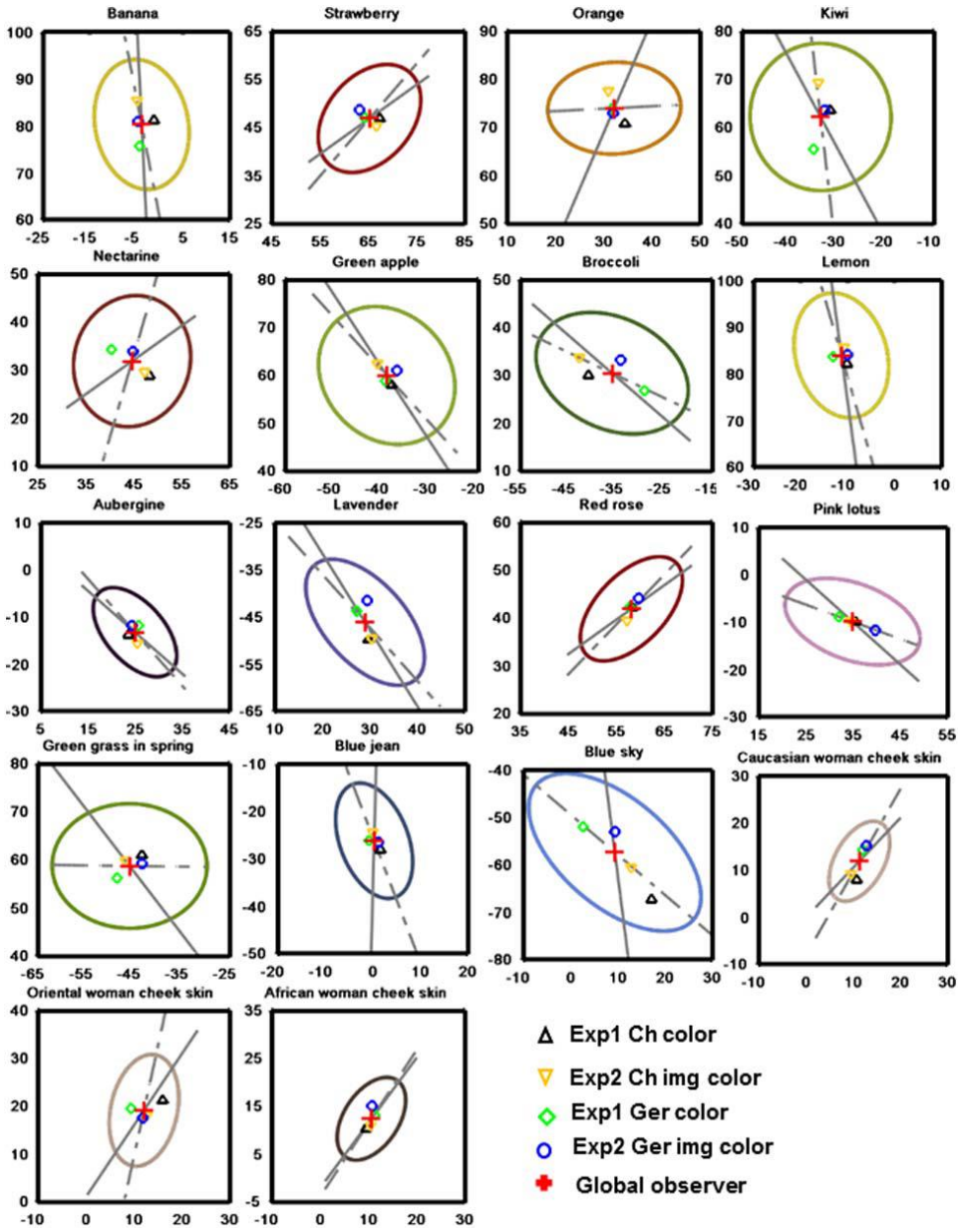


Figure 4.6 – Memory color centers and tolerance boundary ellipses of the average global observer for each of the 18 test objects plotted in the a^*b^* chromaticity plane. The solid grey line indicates the orientation (hue angle) of the memory color center in relation to the coordinate origin, while the dashed-dotted grey line visualizes the orientation of the ellipse's semi-major axis. Please note that the ellipse axes are reduced by a factor of 2. For comparative purposes, the memory color centers of the average Chinese and German observer obatined from both experiments are also indicated. Figures were taken from Ref. [191].

the respective test object was displayed as an additional clue above the color patch and below the object's name. For the second part, 31 Chinese and 25 German observers participated. In Fig. 4.6, the results of both memory color matching experiments are summarized. As can be seen, for each familiar test object the regional memory color centers of both experimental parts are all located within the corresponding matching tolerance ellipses assigned to the global observer obtained by pooling the Chinese and German data. In fact, the overlap of the regional tolerance ellipses is much larger than the extent of the differences between the respective memory color centers. In addition, the typically observed inter-observer variability as an inherent feature of the memory color matching is more than twice as large as the overall

difference between the regional memory color centers. These results in principle confirm the conclusion drawn by Smet *et al.* [55] who stated that the impact of the cultural background on the assessment of memory colors was not significant compared to the relatively large effect of inter-observer variability.

The four different studies discussed above basically give a good overview of how display-based experiments on the assessment of memory colors can be devised and what kind of results could be expected. Even though some of these studies, in particular the ones that make use of Gaussian modeling, provide a conceptionally excellent approach regarding the way color appearance ratings of familiar objects are collected and analyzed, all of them failed to offer suitable viewing and adaptation conditions with respect to real lighting applications – a problem which in a similar manner may also be reported for the studies based on abstract Munsell color samples or the ones using a viewing cabinet approach. In each case, the experiments were performed in a dim/dark environment with the self-luminous display being more or less the only source of light controlling the observers' state of adaptation. Hence, as stated by Tarczali *et al.* [41], the reported results are limited to the very specific viewing conditions of the respective study. For applications such as digital image color enhancement or in the post-production of the movie industry, where the tweaking of some image content is performed in similar environmental conditions, a color appearance model like the one previously discussed could be used to deal with another, slightly different viewing situation so that the reported display-based memory colors and their corresponding tolerance boundaries could still be used for image enhancement. As stated by Bodrogi *et al.* [190], a possible strategy for such applications would be for example to map the color gamut of a certain memory-related image content onto its respective tolerance volume in a three-dimensional color space. However, in case of illuminating real objects for example in shops, supermarkets, or the retail industry in general as well as in case of office and home lighting applications, the differences in viewing and adaptation conditions between such applications and the display-based experiments are far too big to be still adequately absorbed by any available color appearance model.

Besides providing a much more immersive, three-dimensional viewing experience, which display-based methods usually could not, real lighting applications also include the proper perception of object features such as texture, size, and shape which evidently have a significant influence on color constancy [187, 188]. Hence, it is most likely, that similar effects would also be observed for memory colors which, at least when using computer monitors or other two-dimensional output devices, could not appropriately be mirrored in a display-based experiment. Furthermore, from a perceptual point of view, observers are in a completely different, non-comparable cognitive state when being exposed to a real life lighting situation compared to what is expected when they only assess object representations and images or even worse abstract color patches while being seated in front of a computer monitor.

Due to all these severe drawbacks mentioned above, it is very questionable if the results and the knowledge gained from this kind of experiments could be applied to real lighting applications. Strictly speaking, the "simplifying" viewing cabinet approaches adopted by Sanders (see Sec. 3.2.1) and Smet *et al.* (see Sec. 3.2.4) must also be ranked somewhere between these display-based methods and real lighting applications and as such cannot provide the required complexity regarding a comprehensive description of the latter. For this apparent reason, the investigations reported in the subsequent sections of this thesis appear to be justified and necessary in order to develop a proper notion of how the mechanisms of the concept of memory colors work in the context of realistic viewing and adaptation conditions.

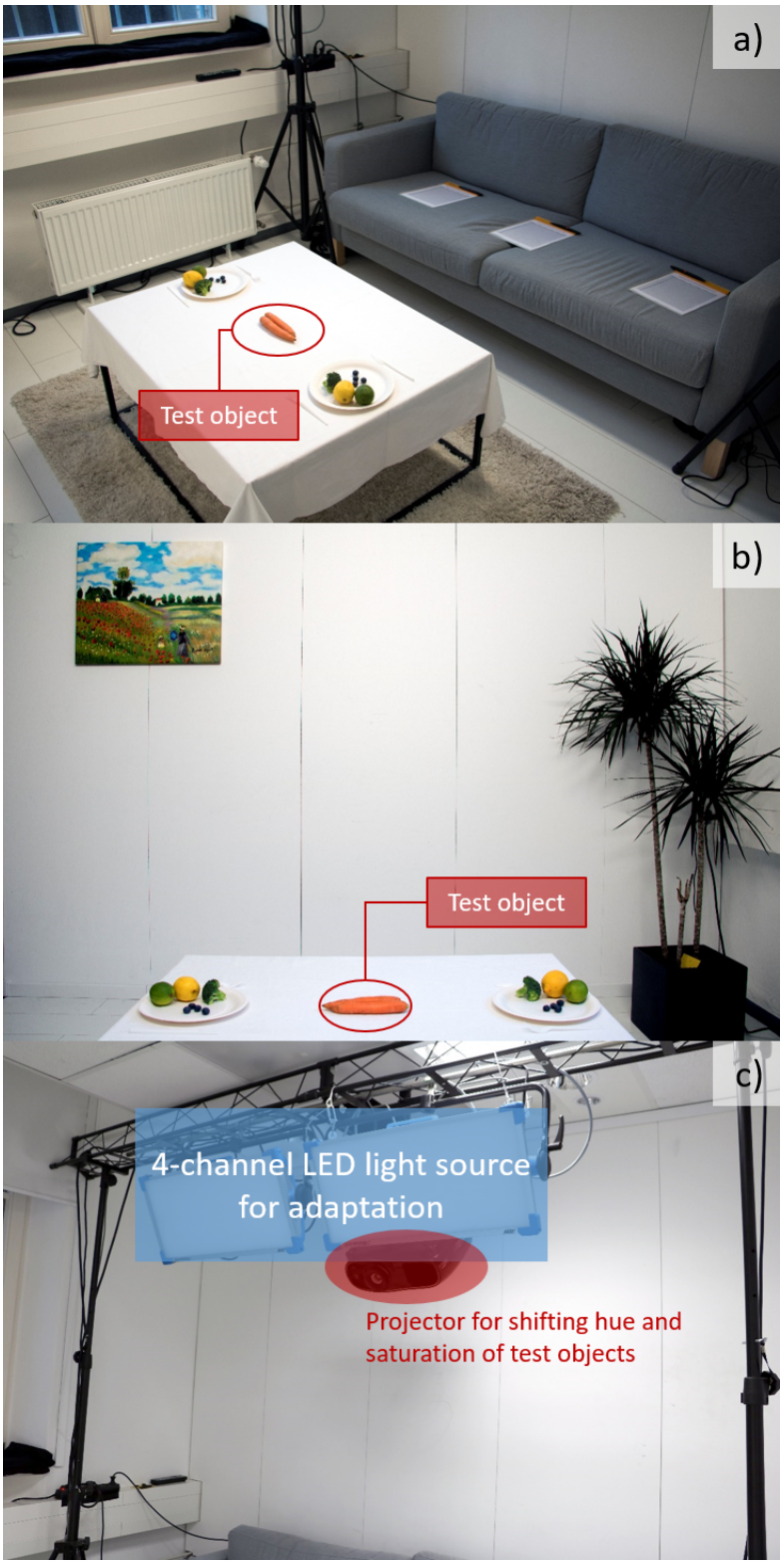


Figure 4.7 – Experimental setup for investigating the impact of long-term memory on the color appearance ratings of familiar test objects. From top to bottom: a) Overview of the experimental arrangement, b) approximate representation of the observers' perspective when assessing the objects, and c) assembly of the LCD projector and the four-channel LED light source. Please note that only the test object should be assessed by the observers, while the additional colored objects being present in the experimental setup were intended to create a specific context to trigger the cognitive mechanisms of color constancy for quickly reaching a stable chromatic adaptation.

4.2 MEMORY COLORS UNDER REALISTIC VIEWING CONDITIONS

With the deliberations of the previous section it should have become clear that for the conception of this thesis, it was not an option to perform another display-based experiment regarding the assessment and perception of memory colors, even if for sure further improvements to the reported studies might still be feasible. However, with the main focus of the current work being on real lighting scenarios and due to the lack of appropriate publications in the literature, a new experimental methodology should be developed whose results are more relevant for lighting applications than those of previous studies. Further details will be given in the remaining parts of this chapter.

4.2.1 *Shortcomings of Previous Studies and Open Research Questions*

In advance of having a closer look at the newly devised experimental setup for the evaluation of the characteristics of memory colors, this section is intended to first give an overview of the open research questions that should be addressed throughout this thesis. As mentioned previously, severe shortcomings in the conceptional design and/or the way observer data are treated can be identified not just for the display-based methods but also for the somewhat more sophisticated viewing cabinet approaches of Sanders (see Sec. 3.2.1) and Smet *et al.* (see Sec. 3.2.4). Due to those shortcomings, which should be summarized in the following, it is very doubtful that the results obtained and the conclusions drawn from these experiments can be applied to real lighting situations in a comprehensive manner.

With both experiments, similar to the display-based approaches, lacking realistic viewing and adaptation conditions, the reported results must actually be considered as being limited to the very specific viewing conditions of the respective study. In both cases, the test objects could only be observed through a relatively small quadratic aperture in the front panel of the viewing cabinet (see Figs. 3.3 and 3.7), which is far from being a realistic lighting situation. Instead of providing an immersive visual experience, which is something only experiments performed in a real-sized room can offer, the test objects of both studies were presented to the observers in a very abstracted fashion isolated from any kind of context which, however, would have been required for drawing universally valid conclusions.

Another point of criticism regarding both studies is the more or less arbitrary and very arguable test object selection. With the test objects of hand, face, tea, butter, beefsteak, and potato chips being selected by Sanders and of green apple, ripe banana, orange, dried lavender, smurf®, strawberry yogurt, sliced cucumber, cauliflower, Caucasian skin, and a neutral gray sphere being chosen by Smet *at al.*, the question arises to which extent these objects represent typical memory colors. Based on a thorough analysis of the corresponding research papers, there is no reason to believe that these sets of test objects are suitable selections for the evaluation of the characteristics of memory colors, which must therefore be doubted. Especially the inclusion of test objects such as tea, butter, potato chips, strawberry yogurt or a smurf® appears to be slightly dubious. Furthermore, it should be noted that the object selection of both studies lacks of hue coverage regarding the more saturated parts of the hue circle. This obviously decribes only a smaller part of the whole picture (characteristics of memory colors in the saturated parts of the color space might differ significantly from those being of lower saturation) and, as stated by Davis *et al.* [19, 194], could be problematic for the predictive performance of a color quality metric based on such low to medium saturated color samples. In particular, the peaked spectra of white LED light sources could be optimized for a high metric value even though the actual perceived color quality is much poorer.

By the inclusion of higher saturated color samples in the metric definition, the risk of such malpredictions would be reduced to a minimum [153].

Last but not least, both Sanders and Smet *et al.*, from a present-day perspective, made use of very questionable color spaces upon which they built their whole data analysis and argumentation. While Sanders, for historic reasons, adopted the anything but perceptually uniform CIE 1931 (x, y) chromaticity space in conjunction with an outdated and poorly performing translational chromatic adaptation transform, Smet *et al.* employed the little-known and rarely used IPT color space, which was at least extended by the latest CAT02 transform allowing, in contrast to the method used by Sanders, for a more proper modeling of the observers' state of chromatic adaptation. However, as conceded by Smet *et al.* [40], the IPT color space was only chosen because of the presence of a self-luminous background in their experimental design, which prevented the use of a more sophisticated and most likely better suited CIECAM02-based color space.

In order to overcome the above mentioned drawbacks of previous studies on the assessment of memory colors, considerable improvements regarding the conceptional design of a new experimental approach are required and will be presented in the subsequent parts of this thesis. Based on these improvements and, consequently, the way observer data must be treated, the following open research questions can additionally be deduced. First of all, with a new experimental setup being adopted in the current work, one might wonder if the general tendencies of the characteristics of memory colors derived from various other studies [25, 31, 33–35, 37–41] could also be confirmed here. In particular, it is of interest whether or not the typical colors assigned to certain familiar test objects still tend to show increased chroma values and hue angles moving towards the typical hues commonly associated with the respective test object when being recalled from memory.

Furthermore, it should be analyzed throughout this thesis whether or not variations in the adaptation conditions and the cultural background of the observers have a significant impact on the memory color assessments. In this context, one may wonder how large the variability in the color appearance ratings of familiar test objects between the individual observers of the same cultural backround is in relation to the variability observed for the average ratings between different cultural observer groups. If the former is found to be of the same order of magnitude or even larger than the latter, the potential impact of the observers' cultural background on the assessment of memory colors must be concluded to be of no practical importance, which would basically confirm the findings reported in Ref. [55]. Regarding the impact of variations in the adaptation conditions, a similar analysis shall be performed as part of this thesis.

Based on the outcome of the improved experimental approach reported at a later stage of this chapter, the research topic that should be addressed next concerns the development of an updated version of an improved memory-based color quality metric for the color rendition evaluation of white light sources in terms of visual appreciation. Compared to the primal work of Sanders and Smet *et al.* discussed in Secs. 3.2.1 and 3.2.4, respectively, the questions are i) how such an improved memory-based color quality metric should be defined and ii) in which manner a performance testing must be conducted in order to validate the superiority of the new proposal over existing methods and, consequently, to deduce reasons for the rather bad performance or even the failing of the latter.

In this context, individual weighting of the adopted color samples should be a key feature of the newly proposed color quality metric. Bearing in mind that for the visual appreciation of a perceived lighting scene certain colors were shown to be more important than others [20, 195, 196], the inclusion of such additional degrees of freedom is considered to be indispens-

able for achieving superior performance in adequately modeling human color preference for lighting applications. Hence, besides providing a more suitable and well-founded test object selection, the question of how these test samples must be weighted shall also be addressed in the course of this thesis.

As should be noticed from this overview section, the identified problems of previous studies on memory color assessments and the open research questions resulting out of them are indeed manifold. However, before their implications will be discussed in detail and before a new proposal of an improved memory-based color quality metric, which considers both the impact of the different adaptation conditions and the observers' cultural background, can eventually be derived, the newly devised experimental setup and test protocol intended to overcome the drawbacks of the previous studies shall be presented first.

4.2.2 *Experimental Setup*

In order to investigate long-term memory effects regarding the color appearance rating of familiar test objects, a new experimental setup was devised which can be considered as an extension of the "simplifying" viewing cabinet approach used by Sanders [25] and Smet *et al.* [40] to more realistic viewing and adaptation conditions. As can be seen from Fig. 4.7, the experiments were performed in a white painted room being equipped with a neutral grey couch, a beige-white shag carpet with a rectangular 80 cm×100 cm coffee table on it, a small but colorful oil painting and a green indoor palm. The ceiling and floor of the room were also kept in white with approximately the same reflectance factor as the walls (~0.8). The coffee table was covered with a white tablecloth and the test object was centrally placed on top of it together with some additional small, naturally colored objects which were arranged on white paper plates with white plastic cutlery (fork and knife) on both sides.

Figure 4.8 – An example of the masking image for a specific test object (here: banana) which was projected on the table. During the experiments the color of the object mask was changed leading to a hue and/or chroma variation of the respective object assessed by the observers. In order to avoid a mixed adaptation state, the dominantly black part of the image ensured that no additional light (or at least only a negligibly small amount of light) leaked through the projector.

Since chromatic adaptation to the white point of the predominant illumination always involves a cognitive component, these additional colored objects together with the oil painting and the indoor palm supported the human visual system to quickly reach a stable adaptation level by creating a specific context which enhanced the mechanisms of color constancy [197]. Without these objects, one would have excluded the cognitive color aspects of the hu-

man visual system [198] and the scene would have appeared too sterile not resembling a real viewing situation any more. In order to avoid simultaneous color contrast and strong interreflections of chromatic radiation, the additional objects were chosen to be small compared to the white adaptation area in the visual field.

During the experiments, three subjects were tested at the same time. They were seated on the couch and asked to observe the scene arranged on the table. The illumination was realized by the combination of a four-channel LED light source (RGBWW) and an LCD projector pointed towards the table's surface. While the LED light source offering Lambertian emission was used to set the ambient light and adaptation condition, the projector enabled shifting the perceived hues and chroma levels of the test objects over a wide range of varieties. By applying a thorough and careful masking individually designed for each test object, which was projected as an image on the table (see Fig. 4.8), the color appearance of the respective test objects could be changed while keeping the observers' chromatic adaptation state constant. Special care had to be taken to perform the masking and the orientation of the test object in such a way that no disturbingly unnatural shadows or colored halos occurred on the white tablecloth and/or the test object.

In order to allow for a proper comparison with previous studies, the correlated color temperature (CCT) of the ambient light was set to approximately 5600 K with the corresponding spectrum being optimized for excellent color rendition. In Fig. 4.9, the normalized measured spectral radiance of the optimized four-channel LED light source is shown in comparison to the D56 reference illuminant. All in all, a general color fidelity index R_f [152] of 91.7 could be achieved with CIE1964 chromaticity coordinates of $x_{10} = 0.3344$ and $y_{10} = 0.3463$. The luminance measured on a diffuse, almost perfectly white ($\Delta\rho < 0.15\%$) Spectralon® reflectance standard placed at the test objects' position was 654.6 cd m^{-2} whereas the measurement of the luminance of the white tablecloth yielded 534.7 cd m^{-2}, given a corresponding horizontal illuminance of 2030 lx. The uniformity of the illumination, assessed by measuring the

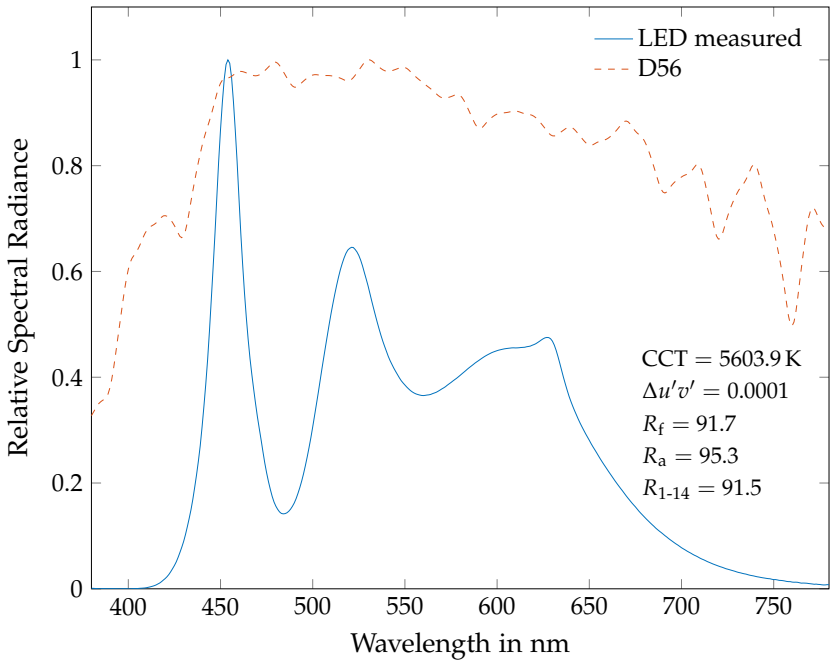


Figure 4.9 – Normalized measured spectral radiance of the optimized four-channel LED light source setting the ambient illumination and adaptation conditions to a CCT of 5600 K. Compared to the D56 reference illuminant, excellent color rendition and white point preservation could be achieved.

horizontal illuminance at several locations across the coffee table, was $E_{\text{min}}/\bar{E} = 0.97$ with $E_{\text{min}}/E_{\text{max}} = 0.94$.

It should be further mentioned that due to the room geometry the light emitting surfaces of the LED light source could be oriented in such away that similar values of illuminance and uniformity were obtained in the central part of the white opposing wall between the oil painting and the plant as seen in Fig. 4.7(b). This area forms the peripheral viewing field of the subjects during the assessment of the test objects. The corresponding illuminance was 1989 lx with a uniformity of $E_{\text{min}}/\bar{E} = 0.83$. Measuring the luminance of the white wall yielded a value of 492.9 cd m^{-2} which was slightly less than the luminance of the white tablecloth but still in an acceptable range. The high illuminance level of approximately 2000 lx of the ambient light in combination with the relatively homogeneously illuminated viewing field (table plus opposing wall) guaranteed a fast, stable, and complete chromatic adaptation of the subjects participating in the experiments (see Refs. [197, 199]).

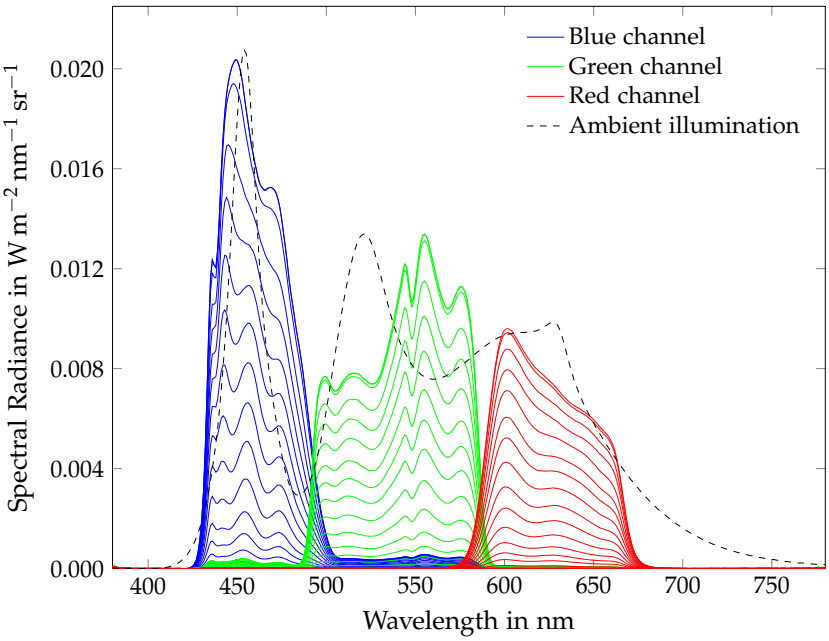


Figure 4.10 – Basis spectra of the red, green and blue color channel of the projection system measured on a Spectralon® target using a Konica Minolta CS-2000 spectroradiometer. The spectral distribution of the optimized ambient illumination was measured in the same manner with a completely black image being projected onto the coffee table.

In addition to knowing the spectral composition of the ambient light, a spectral characterization of the projector was necessary. For this purpose, each color channel (RGB) of the projector was ramped separately from 0 to 255 in increments of 16 (with the LED light source being turned off) while the resulting spectral radiances emitted from the Spectralon® target placed on the coffee table at the test objects' position were measured. Fig. 4.10 summarizes these so-called basis spectra and compares them to the likewise measured spectrum of the ambient illumination when a completely black image was projected onto the table.

Before each of the experimental test sessions, which all took place on separate days, both the four-channel LED light source and the LCD projector were turned on for at least two hours in order to guarantee thermal stability of these two lighting systems throughout the experiments. For this purpose, the LED light source was immediately set to the optimized ambient illumination, whereas a homogeneously light-grey image at approximately 80 % of the peak white was sent to the projector.

Moreover, the repeatability of the illumination settings and the long-term stability of both lighting systems were tested by measuring and comparing the $u'v'$ coordinates of the projector's basis spectra and of the ambient illumination before (after the two hours of warming-up) and after each test session using a Konica Minolta CS-2000 spectroradiometer. The observed color differences $\Delta u'v'$ between the two measurements before and after each session were always smaller than 0.002 for the projector's basis spectra. For the ambient illumination, on the other hand, this short-term color differences were always less than 0.0009 $\Delta u'v'$. Regarding the long-term stability obtained by comparing the measurements between all test sessions, the projector's basis spectra showed deviations of less than 0.004 $\Delta u'v'$, whereas for the ambient illumination deviations of less than 0.0015 $\Delta u'v'$ could be observed.

Based on these measurement results, it can be concluded that both the short- and the long-term stability of the utilized light sources were good enough for collecting precise and accurate color appearance data.

4.2.3 *Object Selection and Stimuli*

In order to achieve a less arbitrary object selection as it might be criticized in previous studies regarding the topic of memory colors, an online survey had been performed in advance of conceiving the experiments whose results should be reported in the following.

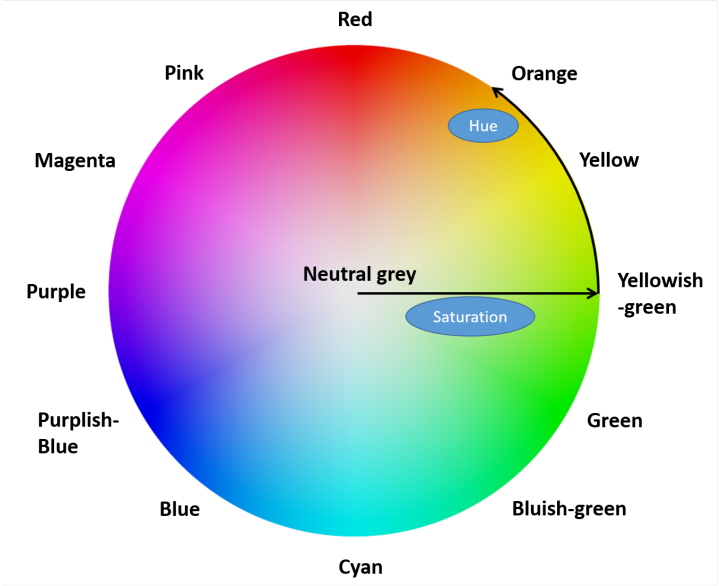


Figure 4.11 – CIELAB hue circle shown to the participants to explain the concepts of hue and saturation. In the subsequent survey, they were asked to name at least one characteristic object for each indicated hue region.

In the survey, which was distributed in German language by e-mail and via the homepage of the Deutsche Lichttechnische Gesellschaft e.V., participants should name the objects that first came into their minds when they thought of a specific color. Before starting with the survey, they received thorough written instructions explaining the characteristics of memory colors and giving a short introduction to the color attributes hue and saturation by using the CIELAB hue circle shown in Fig. 4.11. Regarding the explanation of the concept of memory colors, special care was taken to impart the original definition as devised by Hering [32] who – as stated earlier in this work – used this term to describe the typical color an individual observer has in mind when thinking of or looking at certain familiar objects that he or she

Figure 4.12 – Pie charts representing for each hue region the percentage of participants naming a certain object. Here, the category "Rest" pools both the percentage of all the single objects that were named by only a relatively small number of participants and the percentage of skipped questions.

acquired frequently and memorized stably through his or her experience with the respective objects. To ease the task of naming such objects, participants were advised to think of everyday objects, groceries, plants, etc. that could be found at their homes or were part of their daily routine. Excluded from the survey were only those "objects" that are non-physical, i.e., naming the object "sky" or "water" for the hue region of blue was not allowed.

After reading the instructions, participants could start with the survey whenever they were ready. The task was relatively simple: The color names of Fig. 4.11 defining certain hue regions were consecutively shown to the participants who, for each of these specific color names, were asked to name at least one characteristic object. It was always possible to skip the current color name, if one was not able to name any object at all. No time limit was

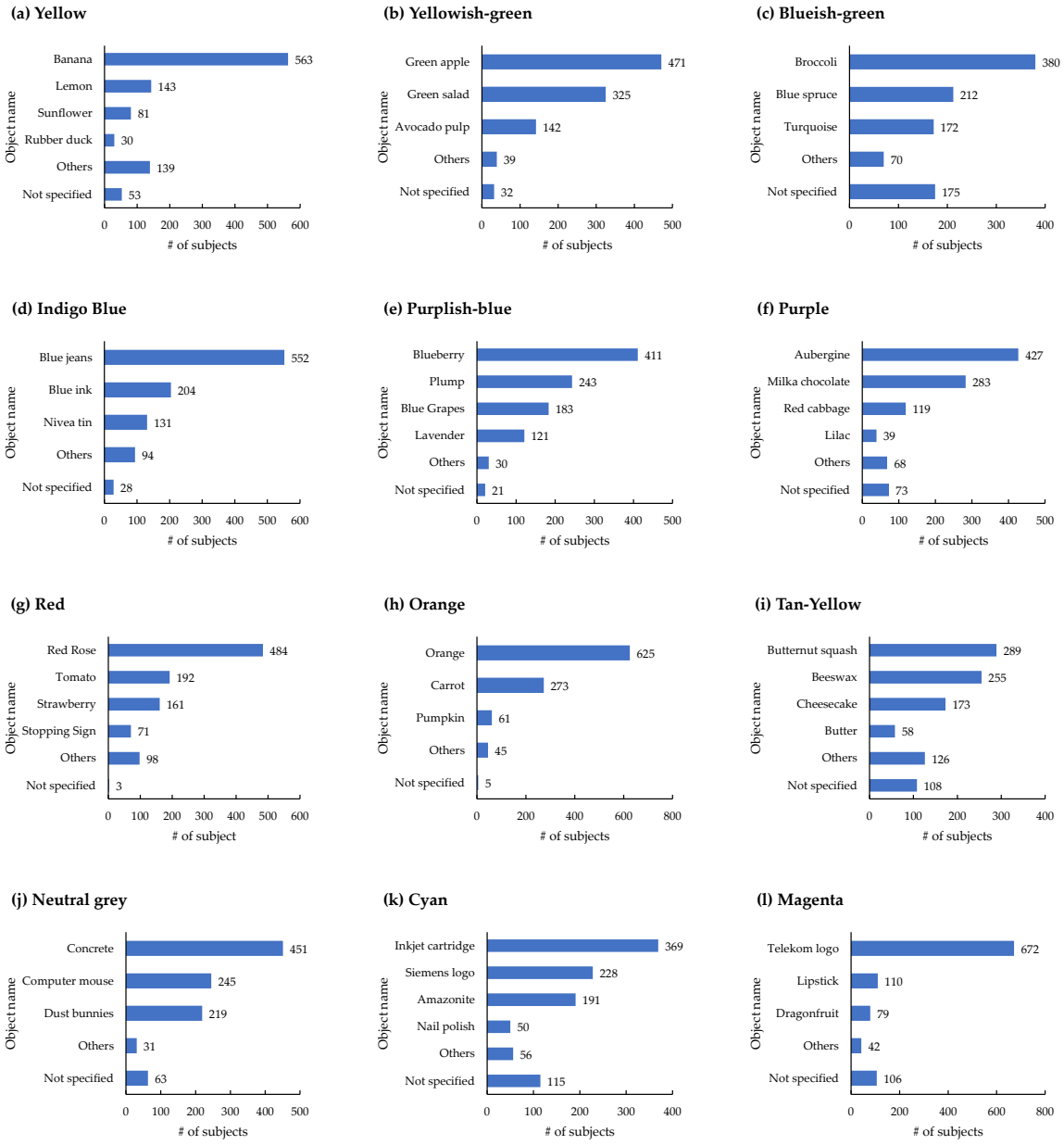


Figure 4.13 – Bar charts representing for each hue region the absolute frequency of responses for a certain object. Here, the category "Others" pools all the single objects that were named by only a relatively small number of participants whereas the category "Not specified" counts the total number of skipped questions.

given to finish the survey. The answers of the participants, which were stored in a database, were subsequently analyzed. The corresponding results are visualized in Figs. 4.12(a)-(l) and 4.13(a)-(l), where, for each hue region, the former show the percentage of participants naming a certain object while the latter give an overview of the absolute frequency of responses.

In total, 1009 subjects took part in this survey. From the analysis of their answers it can be concluded that for some parts of the hue circle (e.g., yellow, red, orange, etc.) there was no problem at all to find a various number of different characteristic objects, while for other parts (e.g., magenta, blueish-green, cyan, etc.), on the other hand, it seemed to be more difficult for the participants to name such typical objects and if they did, their answers were mostly describing artificial or printed items (e.g., the Siemens logo for cyan). Consequently,

these hue regions were excluded from the present study since due to the non-existence of naturally colored objects in these areas they seem to play only a minor role for the investigation of long-term memory color effects. Nevertheless, it is interesting to note that the most frequently named characteristic object of all hue regions was the logo of the German Telekom for magenta which is most likely due to the omnipresence of the corresponding "Magenta One"-brand in German advertising and television. A similar argumentation holds for the second most frequently named object, which is the fruit of orange for the color of orange. In both languages English and German, the names of the fruit and of the color are identical which obviously induces an immediate association in people's brains.

Based on the relative frequency of how often a certain object was associated with a specific hue region, a naturally occurring memory color hierarchy could be established and the following familiar test objects were eventually chosen for performing the experiments: banana (yellow, named by 56 % of the subjects), green salad (yellowish-green, named by 32 % of the subjects), broccoli (bluish-green, named by 38 % of the subjects), blue jeans (indigo blue, named by 55 % of the subjects), blue berry (purplish-blue, named by 41 % of the subjects), red cabbage (purple, named by 12 % of the subjects), red rose (red, named by 48 % of the subjects), carrot (orange, named by 27 % of the subjects), butternut squash (tan-yellow, named by 29 % of the subjects), and concrete flowerpot (neutral grey, named by 45 % of the subjects).

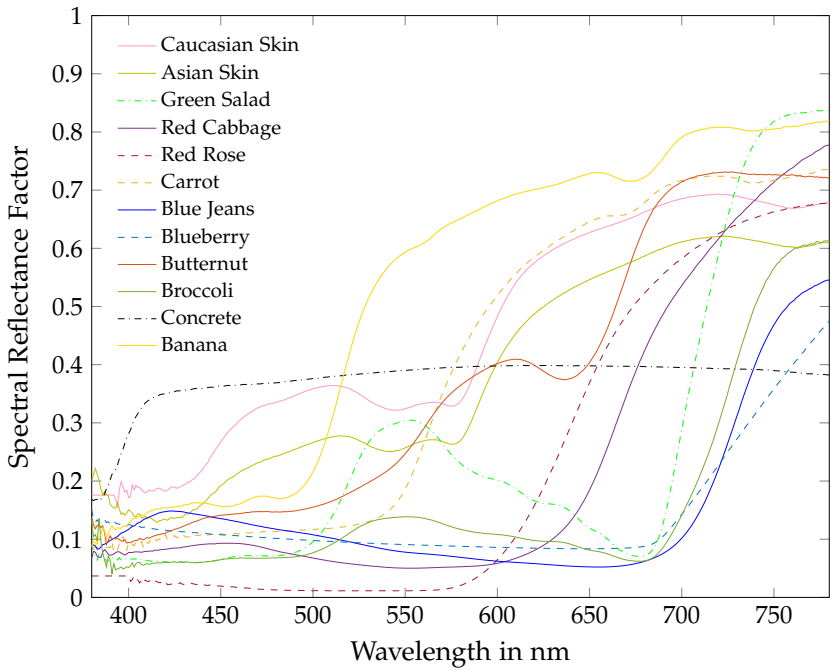


Figure 4.14 – Mean spectral reflectance curves of the twelve familiar test objects. Making use of a temporally stable halogen light source, object measurements were performed using a Konica Minolta CS-2000 spectroradiometer in combination with a Spectralon® reflectance standard.

It should be noted that for the specific hue regions of orange, yellowish-green, and purple the finally selected test objects were not the most frequently named ones in the online survey – actually the objects of orange (named by 62 % of the subjects), green apple (named by 47 % of the subjects), and aubergine (named by 42 % of the subjects) should have been considered in first place. However, pre-tests performed with these objects revealed that their surface characteristics are not Lambertian enough so that, instead of being diffusely scattered for a homogeneous illumination of the test object, the light emitted from the projector and sent upon the object's surface caused to much specular reflections into the observer's eye. For this

reason, it was decided to choose the second or even third (in the case of red cabbage) most named object, which did not show such perturbations on its surface, for the final experiments.

In addition to the ten previously mentioned familiar objects, two different kinds of skin colors were also added for testing, which was driven by the thought that people in their daily routine are usually more often confronted with skin colors than with any other of the chosen memory color objects. Since most people spend most of their life time in the company of others, skin colors are always present to their perception – except when they sleep – and, therefore, are best suited for being indelibly impressed on people's long-term memory which is supposed to lead to relatively consistent ratings among different observers. Hence, skin colors in general fulfill all requirements imposed on a typical memory color with respect to its original definition [32] and, consequently, should be considered in the experiments.

Using exactly the same experimental setup as for the "standard" test objects described in the previous section, the skin colors were presented to the observers in the form of the backside of the right hand of real human models. During the experiments, the hand models were asked to keep their hand lying still on the coffee table so that observers could easily rate its changing color appearance caused by the emitted light of the projector without being distracted by any kind of movements. For the present study, Asian and Caucasian skin color were selected with the respective models being recruited from our faculty staff. Other ethnic skin colors, on the other hand, were excluded from the study simply because of the fact that at the University of Darmstadt it was impossible to find suitable hand models who were willing to sit still for several hours in front of a group of strangers/subjects rating their skin color appearance.

Hence, a total number of twelve different familiar test objects was finally selected for performing the experiments. Their corresponding spectral reflectance curves were calculated from spectral radiance measurements of the objects' surfaces which were homogeneously

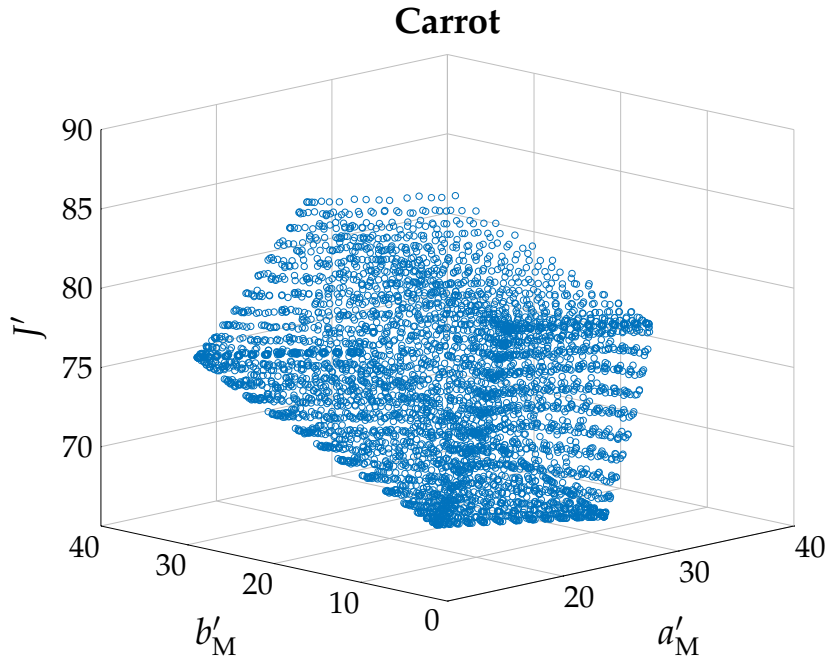


Figure 4.15 – Color gamut of the test object of carrot representing its variations in color appearance in CIECAM02-UCS. Each grid point is obtained by a certain combination of the projector's basis spectra illuminating the test object and can therefore easily be expressed in terms of device-dependent projector RGB values.

Table 4.1 – Mean values and corresponding $\pm 1\sigma$-intervals of the luminance measure and the CIECAM02-UCS lightness parameter for each test object.

Test Object	Mean Luminance $\overline{L}$ in cd m^{-2}	Mean Lightness Parameter $\overline{J'}$
Asian Skin	294.69 ± 5.92	76.17 ± 0.51
Banana	499.09 ± 7.23	91.03 ± 0.57
Blueberry	91.23 ± 2.75	46.91 ± 0.52
Blue Jeans	77.94 ± 2.22	43.85 ± 0.52
Broccoli	107.86 ± 3.06	49.78 ± 0.54
Butternut Squash	266.08 ± 7.34	73.27 ± 0.51
Carrot	290.98 ± 6.38	76.31 ± 0.46
Caucasian Skin	373.73 ± 8.12	83.09 ± 0.55
Concrete Flowerpot	379.54 ± 7.78	83.17 ± 0.54
Green Salad	202.91 ± 5.49	64.36 ± 0.53
Red Cabbage	64.82 ± 2.56	40.79 ± 0.58
Red Rose	36.45 ± 0.79	32.14 ± 0.28

illuminated with a temporally stable halogen light source. These measurements were conducted using again the Konica Minolta CS-2000 spectroradiometer in combination with the Spectralon® reflectance standard. For each test object, several measurements of characteristic surface points were averaged to come up with the finally used spectral reflectance curves shown in Fig. 4.14. The respective colorimetric data computed under the assumption of reference illumination are tabulated in Sec. A.1 of the Appendix.

From the spectral reflectances it was possible to calculate the CIECAM02-UCS color coordinates $J'a'_{\mathrm{M}}b'_{\mathrm{M}}$ [97] mathematically describing the color appearance of the test objects rendered under each combination of the spectral power distributions (SPDs) emitted by the projector (see Fig. 4.10) and the SPD of the optimized four-channel LED light source (see Fig. 4.9). Here, complete adaptation of the observers to the white point of the ambient illumination was assumed. Furthermore, the 10° color matching functions were used in combination with the same parameters as recommended by the CIE for the calculation of the R_{f} color fidelity index [152]. These parameters are the relative luminance of the background $Y_{\mathrm{b}} = 20$, the degree of chromatic adaptation $D = 1$ ($\widehat{=}$ complete adaptation), and the luminance of the adapting field $L_{\mathrm{A}} = 100\,\mathrm{cd\,m^{-2}}$. Furthermore, an average surround was assumed giving an exponential non-linearity of $c = 0.69$ and a chromatic surround induction factor of $N_c = 1$.

By plotting the calculated color coordinates in the CIECAM02-based color space, the color gamut of each test object representing its variations in color appearance could be visualized (see Fig. 4.15 for illustration). From these three-dimensional representations, an equal lightness plane corresponding to the largest chromaticity gamut was extracted for each test object to eliminate the influence of different lightness levels on the observers' chromaticity ratings [40]. However, it should be noted that at the same time the object's luminance had to be kept smaller than the luminance value of the white tablecloth the object was lying on in order to avoid the impression of observing a self-luminous object. This additional constraint lead to a reduced chromaticity gamut for some of the test objects, especially for those of light colors (e.g., Asian and Caucasian skin, concrete flowerpot,...), compared to what theoretically would have been possible. The resulting luminance and CIECAM02-UCS lightness values are summarized in Table 4.1.

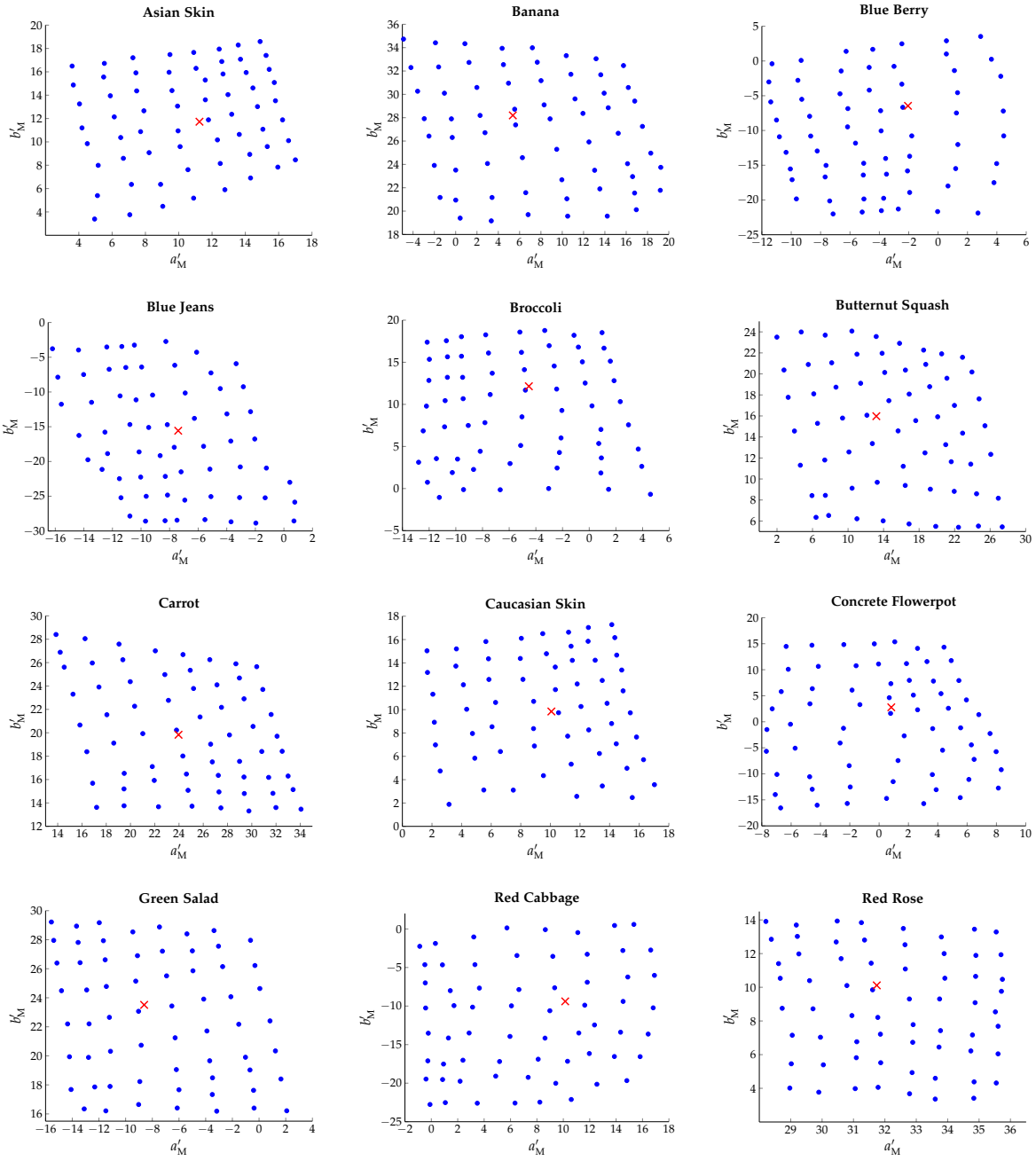


Figure 4.16 – CIECAM02-UCS chromaticities (blue dots) for each of the twelve familiar test objects used for testing the assessment of memory colors at 5600 K ambient illumination. Red crosses mark the objects' chromaticities as they would appear under the 5600 K LED ambient illumination only.

Here, it should be again emphasized that each grid point lying within the chromaticity gamut of a specific test object is the result of a certain combination of the projector's basis spectra illuminating the object as described above and can therefore be directly related to the RGB triplets with which the projector must be driven to obtain this specific combination of basis spectra. Hence, in order to finally end up with an approximately uniform chromaticity grid at a given lightness, inverse mapping could be used to determine for each test object a set of corresponding projector RGB values resulting in the desired object chromaticites, which are shown in Fig. 4.16 for the individual test objects.

During the experiments, these sets of projector RGB values were used to illuminate the test objects which change their color appearance accordingly. The order in which the different resulting spectra and, therefore, the object chromaticities were presented to the observers was completely randomized in order to average out potential bias effects. Following the test protocol of Smet *et al.* [40], 20 additional RGB triplets were randomly selected from the set of RGB values for each test object and added to the corresponding test sequence in order to provide a possibility to check for consistency in the observers' memory color appearance ratings, i.e., for each test object 20 randomly chosen chromaticities were rated twice within a single test sequence. Furthermore, a small training set of 10 RGB triplets was provided for the observers to get familiar with the rating scale and test procedure. Hence, the whole test sequence for each of the familiar test objects consisted of approximately 95 different illumination conditions.

4.2.4 *Memory Color Assessments and Experimental Results*

In total, 26 male and 18 female observers partcipiated in the experiments. Most of them were recruited among the university students and faculty staff showing a varying degree of expertise in color science with ages ranging between 19 and 64 years (∅ 26.5). Quantifying this level of expertise was basically based on a self-assessment of the observers complemented by what they stated as their occupation, which ranged from undergraduate students of various disciplines with no or little overlap with color science to graduate students and faculty staff from our institute which, however, were not necessarily working in the field of color research. With no distinction being made between observers regarding the age and experience aspect in the evaluation presented later in this section, the obtained results can be assumed to represent an approximated cross-section of a larger, usually heterogeneous (with respect to the aspects of age and experience in color science) population.

All observers were native Germans which had not been born or grown up abroad in order to avoid unwanted cultural bias which could have potentially been induced when mixing the observer group with participants having a different cultural or ethnical background. Furthermore, all participants had normal or corrected-to-normal color vision which was tested using the Ishihara Test for Colour Deficiency [200], the Standard Pseudoisochromatic Plates Part II for Acquired Color Vision Defects by Ichikawa *et al.* [201], and the Farnsworth-Munsell D-15 Color Vision Test [202, 203].

Before starting the color appearance ratings, observers received thorough oral instructions regarding the experimental procedure. Following Refs. [25, 40], they were asked to rate the color appearance of the currently presented familiar test object according to their preference of how they thought the respective object should look like in reality on a semi-semantic five-level scale of the Likert-type, where "1" means very bad, "2" represents bad, "3" indicates a neutral rating, "4" means good, and "5" stands for very good. Observers were further advised not to stare too long at the test object but rather rate its color appearance intuitively. In order to reduce the influence of the previous color representation on the subsequent appearance rating to a negligible minimum, the participants were additionally asked to look at the white of the surrounding tablecloth or the opposing wall while the emitted projector spectrum was changed, which was always announced by an acoustic signal. In combination with the small nature of the test objects compared to the dimensions of the experimental room and the carefully prepared object masking, this procedure guaranteed the maintenance of a constant chromatic adaptation level of the observers during the whole test sequence.

Each familiar test object was rated by a total number of 15 observers with an approximately equally balanced male-female ratio (due to organizational and scheduling conflicts not all participants were able to rate all test objects). Since three observers could be tested at the same time, five test runs were necessary for each test object. Including the instruction and the pre-test to familiarize with both the rating scale and the test procedure, it took the observers approximately 20 minutes to finish a test run, which was supposed to be short enough to prevent fatigue. During the experiments, each object chromaticity was shown to the observers for exactly 12 seconds before switching to the next one. Within the first 4 seconds they were asked to mentally focus on the new color appearance of the presented test object in order to be able to compare it with their idea of how the respective object should actually look like. After another acoustic signal was given, the remaining 8 seconds could be used by the observers to perform their rating.

The experiments were efficiently organized in such a way that all five test runs for a specific familiar test object were conducted in consecutive order so that the complete testing of this object could be finished within 2.5 hours. During this relatively short period of time, any kind of food alteration or degradation which might have had a negative impact on the experimental outcome could be precluded. This was confirmed by a visual inspection of corresponding high-quality image pairs showing details of each test object before and after the five test runs. In all cases no perceivable differences were observed. The images were always captured under the same fixed illumination (four-channel LED light source optimized to D56) and viewing conditions (same geometry under which the subjects had to perform the experiments) using a Canon 750D digital camera. In addition, no more than one object per day was tested. Hence, with experiments taking place twice a week, all experiments comprising the twelve familiar test objects could be finished within a six-weeks period. The obtained results will be reported in the following.

4.2.4.1 *Observer Variability*

In order to quantify the precision with which the memory color appearance rating was performed and to obtain a measure for the repeatability of the corresponding experiments, inter- and intra-observer variability should initially be analyzed before modeling the observer ratings. For this purpose, two different statistical measures were found to be appropriate. The first one is the so-called PF/3 performance factor, which was first introduced by Guan and Luo [204] to describe the accordance between two sets of congeneric data, while the second one is the standardized residual sum of squares (STRESS) originally developed for multidimensional scaling techniques [205, 206]. Being based on the calculation of the sum of squared deviations between corresponding pairs of data, the STRESS value is somewhat simpler in its definition than the PF/3 measure, which is essentially a combination of three other, suitably weighted statistical measures to one single formula given by

$$\mathrm{PF/3} = 100 \cdot \left((\gamma - 1) + V_{AB} + CV/100 \right) /3, \tag{4.1}$$

where the γ and CV values were proposed by Coates *et al.* [207, 208] and the V_{AB} metric was derived by Schultze [209] in order to perform similar comparison tasks.

Even though STRESS and PF/3 arise from conceptually different definitions, they can both be used as an indicator for the agreement or disagreement between two sets of data in a quite consistent manner. Generally speaking, the smaller these measures are, the better the sample data conforms to the target data, where, in both cases, a value of zero indicates perfect agreement. Regarding their use in color science, they have widely been applied in experimental studies on color discrimination and color-difference thresholds [204, 210–220]

to (i) compare the performance of various color-difference formulae and (ii) to assess inter- and intra-observer variability. As aforementioned, the main focus of this section should be on the latter.

By assuming that, in the present study, each of the n randomly presented chromaticities of a specific test object were rated independently by k observers randomly selected from a larger population, the individual inter-observer performance factor $\mathrm{PF}/3_j$ of observer j with $j = 1, ..., k$ can be calculated by using the following equations:

$$CV_j = 100 \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i - f_j Y_{i,j} \right)^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^{-1}, \tag{4.2}$$

$$\log_{10} \left(\gamma_j \right) = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\log_{10} \left(X_i Y_{i,j}^{-1} \right) - \frac{1}{n} \sum_{i=1}^{n} \log_{10} \left(X_i Y_{i,j}^{-1} \right) \right)^2 \right)^{\frac{1}{2}}, \tag{4.3}$$

and

$$V_{AB,j} = \left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i - F_j Y_{i,j} \right)^2 \left(X_i F_j Y_{i,j} \right)^{-1} \right)^{\frac{1}{2}}, \tag{4.4}$$

with

$$f_j = \left(\sum_{i=1}^{n} X_i Y_{i,j} \right) \left(\sum_{i=1}^{n} Y_{i,j}^2 \right)^{-1}, \tag{4.5}$$

and

$$F_j = \left(\left(\sum_{i=1}^{n} X_i Y_{i,j}^{-1} \right) \left(\sum_{i=1}^{n} X_i^{-1} Y_{i,j} \right)^{-1} \right)^{\frac{1}{2}}, \tag{4.6}$$

where $X_i = \frac{1}{k} \sum_{j=1}^{k} Y_{i,j}$ is the mean observer rating for the object chromaticity i with $i = 1, ..., n$ and $Y_{i,j}$ is the corresponding individual rating of observer j. Hence, we have

$$\mathrm{PF}/3_j = 100 \cdot \left(\left(\gamma_j - 1 \right) + V_{AB,j} + CV_j/100 \right) /3, \tag{4.7}$$

giving the inter-observer performance factor of the individual observer. The final inter-observer PF/3 values are eventually calculated by taking the mean of the individual measures obtained for each test object, i.e.,

$$\mathrm{PF}/3 = \frac{1}{k} \sum_{j=1}^{k} \mathrm{PF}/3_j. \tag{4.8}$$

In contrast to the PF/3 performance factor, the inter-observer STRESS measure demands less calculation steps and is in general easier to interpret since it simply accumulates the total squared residuals of all pairs of data (modified by some additional scaling factor). Its basic definition using the same denotation as introduced above reads

$$\mathrm{STRESS} = \frac{1}{k} \sum_{j=1}^{k} \mathrm{STRESS}_j, \tag{4.9}$$

where

$$\mathrm{STRESS}_j = \left(\left(\sum_{i=1}^{n} \left(X_i - f_j Y_{i,j} \right)^2 \right) \left(\sum_{i=1}^{n} X_i^2 \right)^{-1} \right)^{\frac{1}{2}} \tag{4.10}$$

represents the individual inter-observer STRESS value of the j^{th} observer.

The intra-observer PF/3 and STRESS values, on the other hand, are both calculated in a quite similar manner as their inter-observer counterparts with X_i being simply replaced by a new variable $X_{i,j}$ in the above equations. This new variable denotes the average rating of observer j obtained from the two repeated assessments of object chromaticity i. In addition, $Y_{i,j}$ now represents the individual observer rating of the second assessment. With only a subset of object chromaticities being presented twice for each test object, the summation variable n reduces to the value of the size of this subset, which in the present case would be 20.

Hence, for each familiar test object the inter-observer PF/3 performance factor and STRESS value, giving both an estimate for the precision of the memory color assessments, were calculated by comparing the ratings of the individual observers with the mean appearance ratings of all observers. In contrast, the intra-observer PF/3 and STRESS repeatabilitiy estimates were calculated from the subset of repeated trials provided for each test object to check for observers' consistency. Please note that this procedure is consistent with the color discrimination studies [204, 210–212, 220] but differs from the method chosen by Smet *et al.* [40] who calculated the intra-observer PF/3 values from the ratings of observers repeating the experiments on a different day.

In addition to the inter- and intra-observer PF/3 performance factors and STRESS values, a two-way random effects intraclass correlation coefficient $\mathrm{ICC}(2,k)$ as described by Shrout and Fleiss [221] was also calculated. This measure gives an idea about the reliability of the concept of an average observer based on the ratings of a limited number of k individual observers randomly chosen from a larger population. It can be calculated using the following equation:

$$\mathrm{ICC}(2,k) = \frac{\mathrm{BMS} - \mathrm{EMS}}{\mathrm{BMS} + \frac{1}{n}\left(\mathrm{JMS} - \mathrm{EMS}\right)}, \tag{4.11}$$

Table 4.2 – Average inter- and intra-observer PF/3 performance factors and STRESS values calculated from the visual appearance ratings for the twelve familiar test objects at 5600 K adapted white point. In addition, the resulting $\mathrm{ICC}(2,k)$ values are shown together with their 95 % confidence intervals. Please note that for both measures PF/3 and STRESS the corresponding values of inter- and intra-observer variability are tabulated in the same column with the latter given in parenthesis.

Test Object	Inter-(Intra-)observer PF/3	Inter-(Intra-)observer STRESS	$\mathrm{ICC}(2,k)$	$\mathrm{ICC}(2,k)$ Confidence Interval
Asian Skin	36 (20)	0.28 (0.15)	0.9027	[0.8537, 0.9374]
Banana	34 (18)	0.25 (0.15)	0.9491	[0.9283, 0.9658]
Blueberry	35 (19)	0.27 (0.15)	0.9399	[0.9156, 0.9595]
Blue Jeans	36 (17)	0.29 (0.14)	0.9425	[0.9193, 0.9613]
Broccoli	31 (19)	0.24 (0.15)	0.9753	[0.9653, 0.9833]
Butternut Squash	29 (17)	0.23 (0.13)	0.9689	[0.9563, 0.9790]
Carrot	34 (17)	0.25 (0.14)	0.9348	[0.9089, 0.9560]
Caucasian Skin	35 (23)	0.28 (0.14)	0.9441	[0.9203, 0.9628]
Concrete Flowerpot	34 (17)	0.26 (0.14)	0.9569	[0.9374, 0.9716]
Green Salad	38 (22)	0.28 (0.16)	0.9130	[0.8771, 0.9418]
Red Cabbage	35 (19)	0.27 (0.15)	0.9338	[0.9071, 0.9554]
Red Rose	32 (18)	0.24 (0.13)	0.9414	[0.9175, 0.9607]

with

$$\mathrm{BMS} = k \sum_{i=1}^{n} \frac{\left(\frac{1}{k} \sum_{j=1}^{k} Y_{i,j} - \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{i,j} \right)^2}{n-1}, \tag{4.12}$$

$$\mathrm{JMS} = n \sum_{j=1}^{k} \frac{\left(\frac{1}{n} \sum_{i=1}^{n} Y_{i,j} - \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{i,j} \right)^2}{k-1}, \tag{4.13}$$

and

$$\mathrm{EMS} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} \left(Y_{i,j} - \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{i,j} \right)^2 - (k-1)\,\mathrm{JMS} - (n-1)\,\mathrm{BMS}}{(n-1)\,(k-1)}, \tag{4.14}$$

where $Y_{i,j}$ again denotes the rating of the j^{th} observer on the i^{th} object chromaticity with $i = 1, ..., n$ and $j = 1, ..., k$. The closer the $\mathrm{ICC}(2,k)$ value is to 1, the better is the absolute agreement between the observers' ratings and the more reliable is the assumption of an average observer. Finally, Table 4.2 summarizes for each test object the average inter- and intra-observer PF/3 and STRESS estimates as well as the calculated $\mathrm{ICC}(2,k)$ values together with their corresponding 95 % confidence intervals.

As can be seen from Table 4.2, the inter- and intra-observer PF/3 measures range from 29 to 38 and from 17 to 22 with an average value of 34 and 19, respectively. Recalling that the subjects rated the color appearance of the objects against their long-term memory impression, which is supposed to be at least to some extent different for each individual observer, these results are in very good agreement with the performance factors obtained in the color discrimination experiments performed by Guan and Luo [204, 210, 211] and Xu *et al.* [212], where the typical observer precision and repeatability measures were found to be of the order of 30 PF/3 units for the inter- and 20 PF/3 units for the intra-observer variability.

Compared to the results of Smet *et al.* [40], where average inter- and intra-observer PF/3 values of 40 and 23 were reported, slightly higher observer precision and repeatability could be observed in the present experiments. This improved performance might be explained by the more realistic viewing and adaptation conditions and by the application of a more sophisticated method for selecting suitable familiar test objects reducing the error induced by "unknown" or unfamiliar memory colors, i.e., test objects for which observers would have had difficulties to keep a consistently fixed reference in their mind while performing the color appearance ratings.

Furthermore, remarkably small inter- and intra-observer STRESS values ranging from 0.23 to 0.29 and from 0.13 to 0.16 with an average value of 0.26 and 0.14, respectively, are reported in Table 4.2. In comparison to studies on color discrimination, where typical STRESS values were stated to be of the order of 0.32 for the inter- and 0.37 for the intra-observer variability [219, 220], excellent consistency between the ratings of different observers as well as between the repeated trials of the same observer can be concluded here. This basically indicates a high familiarity of the individual observer with the test objects' typical memory color representations and allows for drawing conclusions about the general characteristics of memory colors based on the current experiments.

Regarding the $\mathrm{ICC}(2,k)$ calculations, excellent performance must be deduced from the individual rating data obtained for each familiar test object showing even slightly higher values (≥ 0.90) than those reported by Smet *et al.* [40]. Based on these results, reliable justification is given that an average observer can be postulated in the following by pooling the individual observers' ratings for each test object.

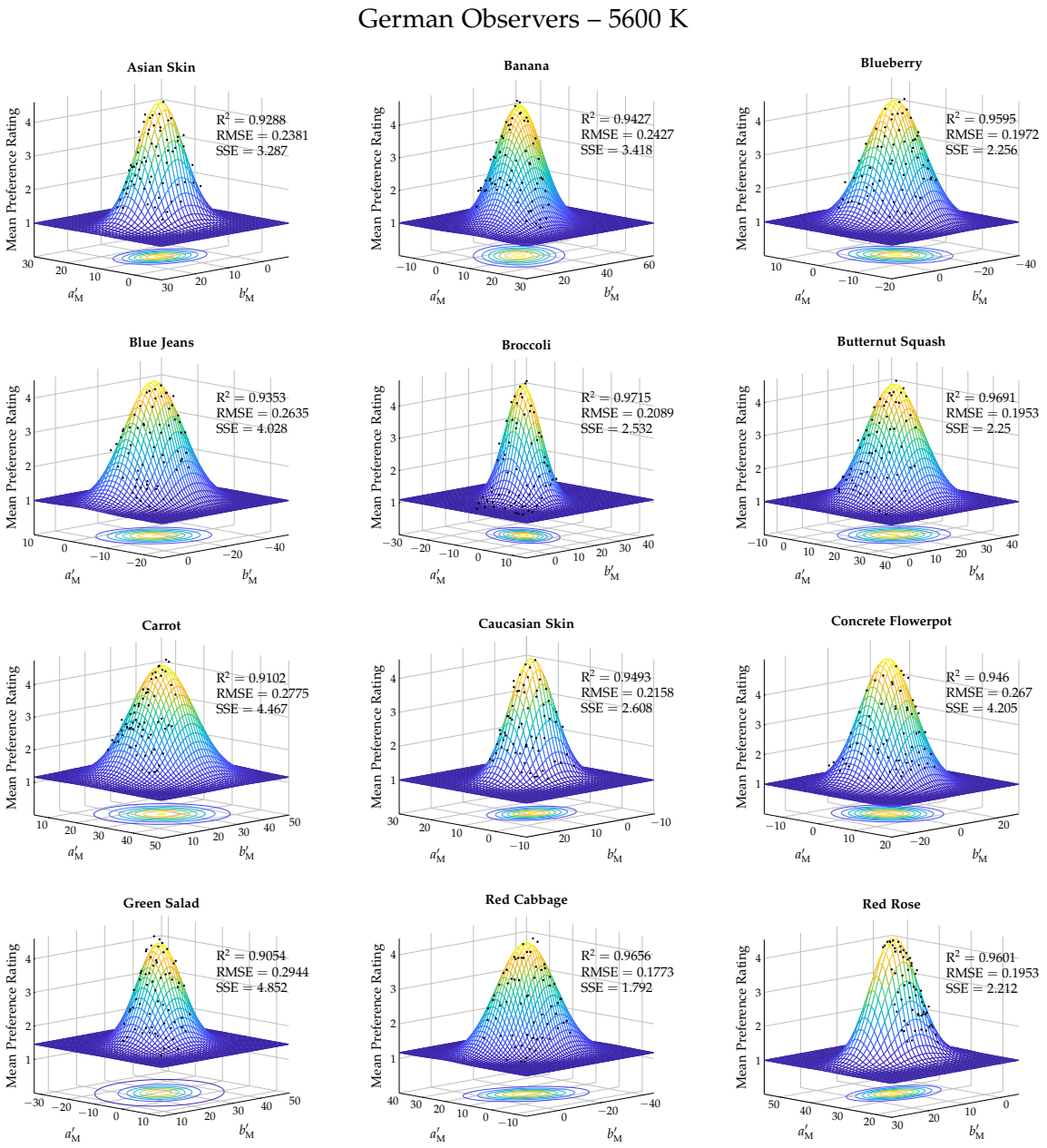


Figure 4.17 – Bivariate Gaussian similarity distributions fitted to the pooled German observer data for each familiar test object modeling the mean preference ratings of an average German observer adapted to the 5600 K ambient illumination in CIECAM02-UCS chromaticity space. Excellent model performance based on the goodness-of-fit statistics can be concluded for all cases. Black dots represent the mean observer ratings for each illumination setting whereas the centroids of the Gaussian distributions define the chromaticity coordinates of the corresponding memory colors.

4.2.4.2 *Gaussian Modeling*

According to the work of Yendrikhovskij *et al.* [38], the most appropriate way of modeling the memory color assessments of the present experiments is to fit for each test object a multivariate Gaussian probability density function to the pooled observer ratings, which besides giving a prototypical representation of the test object's memory color also allows for quantifying the perceived similarity between the respective object's color appearance and the idea of how the object should ideally look like for an average observer.

The model applied here is given by

$$f(\mathbf{x}) = a_1 + a_2 \cdot \exp\left(-\frac{1}{2}\left((\mathbf{x}-\boldsymbol{\mu})^\mathsf{T}\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})\right)\right), \tag{4.15}$$

where $\mathbf{x} = (a'_\mathrm{M}, b'_\mathrm{M})^\mathsf{T}$ gives the object chromaticities a'_M and b'_M of the adopted CIECAM02-based color space, $\boldsymbol{\mu} = (a_3, a_4)^\mathsf{T}$ defines the centroid of the distribution representing the most likely location of the test object's memory color, and $\boldsymbol{\Sigma}$ is the corresponding covariance matrix whose inverse can be expressed by

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} a_5 & a_7 \\ a_7 & a_6 \end{bmatrix}, \tag{4.16}$$

i.e., the multivariate Gaussian model is described by seven parameters $a_1, \ldots, a_7$ which need to be fitted to the pooled observer data.

Table 4.3 – CIECAM02-UCS chromaticity coordinates a'_M and b'_M of the memory colors of the twelve familiar test objects given by the centroids of the fitted multivariate Gaussian probability density functions. For the sake of completeness, the corresponding Pearson correlation coefficients describing the goodness-of-fit of the Gaussian modeling are tabulated in the last column.

Test Object	a'_M	b'_M	Pearson correlation ϱ
Asian Skin	10.29	10.36	0.96
Banana	5.08	32.59	0.97
Blueberry	−3.32	−12.29	0.98
Blue Jeans	−7.79	−17.36	0.97
Broccoli	−8.27	15.38	0.99
Butternut Squash	15.38	20.57	0.98
Carrot	25.64	25.63	0.95
Caucasian Skin	9.88	8.52	0.98
Concrete Flowerpot	0.08	−1.17	0.97
Green Salad	−10.43	26.70	0.95
Red Cabbage	10.56	−14.08	0.98
Red Rose	34.68	11.12	0.98

The corresponding results are illustrated in Fig. 4.17. For each test object, the mean preference rating of an average observer is modeled by a Gaussian distribution in CIECAM02-UCS chromaticity space. In order to visualize the goodness-of-fit, the mean observer ratings for each test setting are shown as black dots. In addition, various goodness-of-fit statistics are reported. These are the basic sum of squared residuals labeled SSE, the adjusted coefficient of determination R^2, and the root-mean-square error denoted by RSME. Furthermore, the Pearson correlation coefficient ϱ between the model predictions and the mean observer ratings were calculated adopting a 5 % significance level.

In all cases, more than 90 % of the total variations in the mean ratings of the pooled data can be explained by the Gaussian modeling. In combination with the relatively small RMSE values (< 0.3) and an average Pearson correlation coefficient $\overline{\varrho}$ of 0.97, an excellent model performance can therefore be concluded. The chromaticity coordinates of the corresponding memory colors given by the centroids of the multivariate Gaussian distributions were finally extracted from each fit and are summarized in Table 4.3 together with the Person correlation coefficients of the respective model.

4.2.4.3 *Characteristics of Memory Colors*

For a further analysis, the chromaticities of the memory color centers reported in Table 4.3 should be compared with the hypothetical chromaticities the test objects would show when being solely illuminated by the CIE D56 reference light source for which the ambient illumination in the present experiments was optimized to define the chromatic adaptation conditions. The expected deviations between both concepts are illustrated in Fig. 4.18.

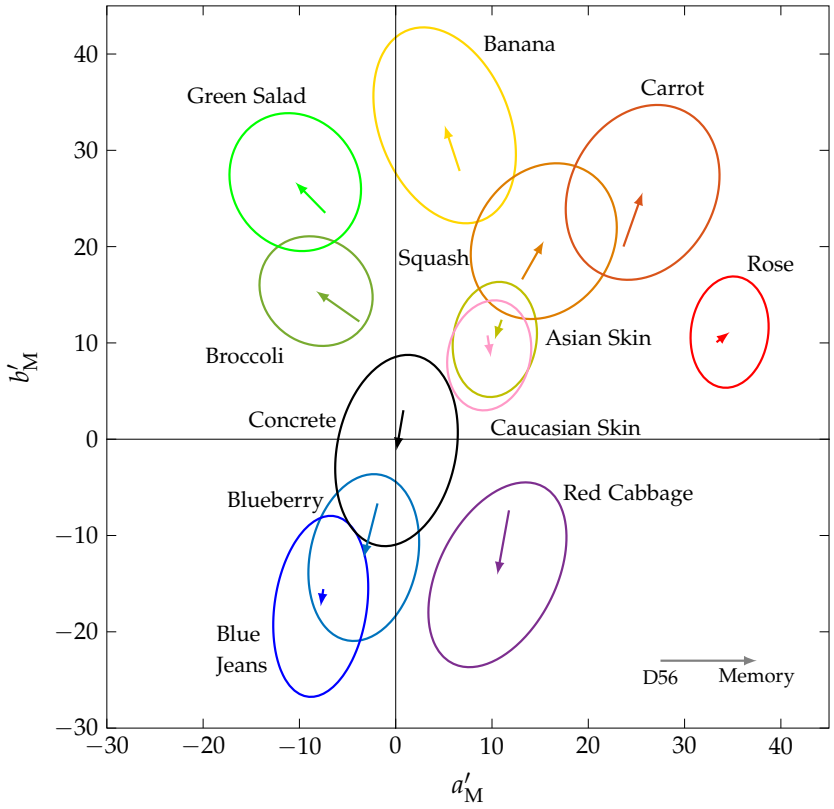


Figure 4.18 – Comparison between the memory color centers and the test object chromaticities rendered using reference illuminant D56. Observed memory-induced chromaticity shifts are indicated by colored arrows. In addition, acceptance boundary ellipses representing an average observer rating of 3 are plotted for each test object.

As can be seen, most of the memory-induced chromaticity shifts indicated by the colored arrows point in the direction of increased CIECAM02-UCS colorfulness M' given as

$$M' = \sqrt{{a'_M}^2 + {b'_M}^2}, \tag{4.17}$$

which according to Ref. [97] is monotonically related to the CIECAM02 perception correlate of colorfulness M by

$$M = \frac{1}{0.0228} \cdot \left(\exp\left(0.0228 \cdot M'\right) - 1\right). \tag{4.18}$$

With the assumption of constant viewing and adaptation conditions being valid in the present experiments, proportionality between the colorfulness M and the CIECAM02 chromatic correlate of chroma C can be postulated yielding $M \propto C$ for each test object. Hence, the memory-induced increase in colorfulness observed from Fig. 4.18 also results in higher chroma correlates for most of the test objects as summarized by Table 4.4 in terms of CIECAM02 chroma increments ΔC, where a plus (minus) sign indicates an increase (decrease) in the

Table 4.4 – Overview of the chromatic differences and of the deviations in the perception correlates of chroma and hue between the memory color centers and the object chromaticities under reference illumination D56.

Test Object	$\Delta E'_{\text{chrom.}}$	ΔC	Δh in °
Asian Skin	2.17	−3.00	−3.23
Banana	4.99	9.30	4.55
Blueberry	5.80	7.69	0.69
Blue Jeans	1.82	2.80	1.62
Broccoli	5.49	6.98	11.11
Butternut Squash	4.57	8.18	1.55
Carrot	5.97	12.03	4.78
Caucasian Skin	2.29	−1.95	−7.73
Concrete Flowerpot	4.24	−2.16	198.83
Green Salad	4.46	7.87	4.06
Red Cabbage	6.80	5.63	−21.01
Red Rose	1.72	3.85	0.95

chroma values of the memory colors in comparison to the chroma values the test objects would show under reference illumination.

From the obtained ΔC values it can be concluded that memory colors of familiar objects generally tend to be shifted towards higher chroma which has also been confirmed by various other studies [25, 31, 33–35, 37–41]. Exceptions hereto can be found for test objects which are naturally of low saturation confirming the findings of Pérez-Carpinell *et al.* [37]. In the present experiments, these are both skin colors as well as the concrete flowerpot. While the former in accordance with the results of Sanders [25] and Smet *et al.* [40] are recalled with a slightly redder hue as they would appear under reference illumination, the latter is shifted towards the neutral origin of the chromaticity diagram resulting in an almost perfect grey color.

Similar shifts in hue are observed for most of the test objects confirming the findings of Newhall *et al.* [34] and Bartleson [35] who concluded that memory colors tend to be shifted towards the typical or dominant hues commonly associated with the respective objects, i.e., the carrot is recalled more orange, the banana is recalled more yellow, etc. as their color appearance under reference illumination D56 would suggest. For each test object, these memory-induced hue shifts Δh were calculated and the corresponding results are also listed in Table 4.4, where the sign convention of the hue angle difference between the memory and the reference representation of the respective object follows the order of the color spectrum from the red to the blue region in a mathematical positive (anticlockwise) sense. This means that a plus sign indicates that the corresponding memory color was found to be slightly more yellow than a red perception, or greener than a yellow perception, or bluer than a green perception, etc., while for the minus sign it is the other way round.

Furthermore, the CIECAM02-UCS chromatic color differences $\Delta E'_{\text{chrom.}}$ defined by the length of the chromaticity arrows between the reference representation of the objects and the corresponding memory colors were calculated giving a measure for the strength of the long-term memory effects on color appearance ratings of familiar objects. On average, these effects are of the order of 4.19 $\Delta E'_{\text{chrom.}}$ (see Table 4.4), where the memory colors of red rose and blue jeans show the smallest deviations with values of 1.72 $\Delta E'_{\text{chrom.}}$ and 1.82 $\Delta E'_{\text{chrom.}}$,

Table 4.5 – Overview of the geometric measures of the chromatic tolerance ellipses shown in Fig. 4.18. Please note that the semi-major axis a and semi-minor axis b are given in units of $\Delta E'_{\text{chrom.}}$.

Test Object	a	b	a/b	θ in °	πab
Asian Skin	6.02	4.32	1.39	80.56	81.79
Banana	10.59	6.78	1.56	111.34	225.52
Blueberry	8.77	5.58	1.57	78.24	153.65
Blue Jeans	9.47	4.78	1.98	81.60	142.39
Broccoli	6.12	5.43	1.13	142.69	104.39
Butternut Squash	8.54	7.08	1.21	55.13	189.83
Carrot	9.48	7.51	1.26	61.92	223.53
Caucasian Skin	5.78	4.27	1.35	76.79	77.64
Concrete Flowerpot	10.05	6.18	1.63	78.81	194.98
Green Salad	7.38	6.58	1.12	122.86	152.59
Red Cabbage	10.32	6.10	1.69	62.98	197.66
Red Rose	5.81	4.02	1.45	82.44	73.39

respectively, while the memory colors of carrot and red cabbage exhibit the largest deviations with values of 5.97 $\Delta E'_{\text{chrom.}}$ and 6.80 $\Delta E'_{\text{chrom.}}$.

In addition to these chromaticity shifts, it is of interest to have a closer look at the shape and orientation of the fitted similarity distributions. For this purpose, contour plots describing the acceptance boundaries of the memory color assessments for each familiar test object were added to Fig. 4.18. These so-called chromatic tolerance ellipses with semi-major axis a and semi-minor axis b given in Table 4.5 were calculated from the corresponding similarity distributions in such a way that their contour line represents an average observer rating value of 3, which is considered to be the just acceptable limit in the present color appearance rating experiments. Object chromaticities that lie outside these contour ellipses are therefore assumed to be unsatisfying for the average observer.

As can be seen from Fig. 4.18, for all familiar test objects the reference D56 chromaticities indicated by the starting points of the chromaticity arrows are located within the acceptance boundaries, but for most test objects they are pretty close to the just acceptable limit given by the respective contour lines, which, from a psychophysical point of view, would lead only to relatively moderate color appearance ratings.

Hence, it can be concluded that reference illumination is not necessarily the best option when long-term memory effects must be considered for the design of high quality light sources. In the absence of an external reference, which is usually the case in all kinds of lighting applications, object colors and, therefore, the color quality of an illumination are always assessed with respect to what people expect the objects to look like in the current scene, i.e., they rate the objects' color appearance against an internal reference which has to be recalled from memory. In order to achieve high preference and user acceptability it is therefore necessary to construct light sources which render the colors of an object in such a way that the obtained chromaticities match its memory color representation, which is obviously not the case for the reference illuminant D56 in the present study.

By comparing the ellipse sizes πab of the different test objects shown in Table 4.5, it can further be stated that both skin colors as well as red rose exhibit the most narrow distributions among all test objects. This indicates that on average observers are more sensitive to changes in the chromaticities of these three objects than to changes in the chromaticities of

any of the remaining objects. At least in the case of the skin colors this might be explained by the fact that most people in their daily routine have frequent visual experiences with human complexions, which is supposed to result in a relatively consistent long-term memory representation even among different observers.

From the shape a/b and orientation θ of the tolerance ellipses shown in Fig. 4.18 and Table 4.5 it can be further concluded that, with the exceptions of red rose and red cabbage, observers tend to be more tolerant of changes in chroma than of shifts in hue direction. Similar results were observed by Smet *et al.* [40, 53] who, as stated previously, used these findings to justify the application of the hue uniform IPT color space [177, 178] in the analysis of their experiments, which, following their argumentation, had to be chosen due to the presence of a self-luminous background in their experimental design. In contrast, it is believed by the author that a CIECAM02-based color space was more appropriate to model the observers' color appearance ratings in the present experiments where no such self-luminous effects need to be considered.

4.2.5 *Comparison with the Results of Smet et al.*

Finally, it is of interest to compare the results obtained in the present experiments with those reported by Smet *et al.* [40, 53]. Unfortunately, this comparison is only possible for the test objects of banana and Caucasian skin, i.e., for objects that were presented to the participants in both studies. The question which should be answered here is whether or not significant differences can be found between the memory color centers of the respective objects as reported in the previous sections in comparison with the findings of Smet *et al.* [40, 53].

In both studies, the test objects were presented to a group of observers in such a way that they changed their color appearance while keeping the observers' chromatic adaptation state constant during the rating process. Apart from the different experimental setups and test objects presented to the observers (see Sec. 3.2.4), a more or less identical test protocol was applied. In general, the viewing cabinet approach slightly reduces the hardware demands and eases the preparation process of the experiments but, obviously, lacks realistic adaptation and viewing conditions. Furthermore, much higher object luminances could be reached in the present work compared to what was reported by Smet *et al.* [40]. Hence, differences in the experimental results of both studies were expected and the implications of the more realistic adaptation and viewing conditions at higher luminance levels achieved by the present experimental setup should be discussed in the following.

For this purpose and in view of a better visualization, Figs. 4.19(a) and 4.19(b) show the contour line plots of the fitted similarity distribution functions of both studies for the test objects of Caucasian skin and banana represented in IPT color space. The transformation of the present results to the IPT color space was performed following the instructions given in the appendix of Ref. [53].

As can be seen, in both cases the contour line plots exhibit similar shape and orientation but distinct differences in size with the results of the present experiments showing much narrower similarity distributions. This means that the observers in the experiments of Smet *et al.* were more tolerant of (or less sensitive to) deviations from the memory color centers defined by the centroids of the distribution functions than the observers in the present experiments, which is most likely due to the much lower object luminances applied there ($82\,\mathrm{cd\,m^{-2}}$ vs. $374\,\mathrm{cd\,m^{-2}}$ for Caucasian skin and $99\,\mathrm{cd\,m^{-2}}$ vs. $499\,\mathrm{cd\,m^{-2}}$ for banana).

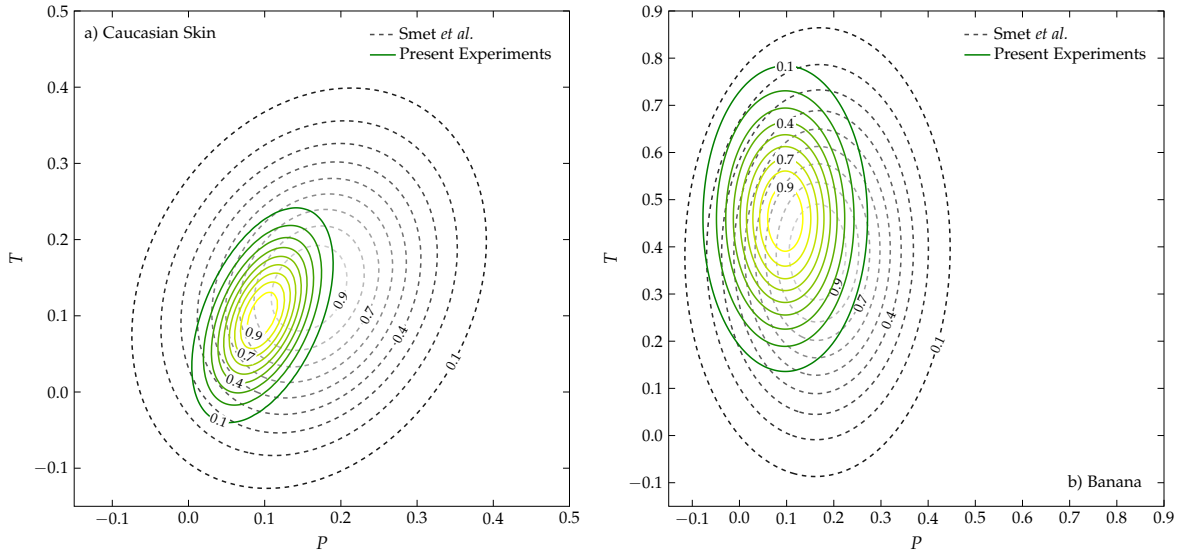


Figure 4.19 – Contour line plots of the fitted, normalized similarity distribution functions obtained in the present experiments (green-to-yellow colormap, solid line) in comparison with the results reported by Smet *et al.* [40, 53] (dark-to-light-gray colormap, dashed line) for the test objects of Caucasian skin (left image) and banana (right image). By applying Box's *M*-test, significant differences can be found in both cases for the covariance matrices defining the shape of the similarity distributions. Furthermore, Hotelling's T^2-test reveals significant differences for the memory color centers reported in the two different studies.

As stated by Pridmore and Melgosa [222] and reported explicitly or implicitly by many other researchers [223–229] performing similar color discrimination experiments, the luminance level in general has a remarkable impact on the size of the tolerance-threshold ellipses. It is found that these ellipses become smaller with increasing luminances allowing for an easier discrimination of small color differences. A comparable effect is observed here for the memory color assessments where the fitted similarity distributions become more narrow at higher object luminances. Hence, small deviations in chromaticity from the "optimal" memory color of an object can more easily be detected in the present experiments due to the higher luminance level, leading to a less preferred rating for the absolute deviations when compared to the results reported by Smet *et al.* [40, 53].

Furthermore, slight differences between the centroids of the similarity distributions of the two different studies can be found for both test objects. In order to determine whether or not these differences are significant, the heteroscedastic version of Hotelling's T^2-test for multivariate, independent data samples [230, 231] was applied assuming a 5 % significance level. Here, heteroscedasticity had to be considered because of the obviously unequal covariance matrices of the data samples that should be compared. Mathematically, the unequality of the covariances was additionally confirmed at a 5 % significance level for both test objects by applying Box's *M*-test [231, 232].

Regarding the outcome of Hotelling's T^2-test, the null hypothesis of equal sample means must be rejected for both test cases. For the test object of banana as well as for the test object of Caucasian skin, a *p*-value smaller than 0.0001 is obtained. Hence, the observed differences between the memory color centers reported by Smet *et al.* [40, 53] and those found in the present experiments for the test objects of banana and Caucasian skin are statistically significant. Nevertheless, it should be pointed out here that even though significant differences can be reported for the memory color centers, it is believed that the significantly different covariance matrices defining the size, shape, and orientation of the corresponding similarity

distribution functions found in the present experiments do have a somewhat greater impact on the final results when being applied for the construction of an updated memory-based color quality metric for the description of the color rendering properties of an arbitrary test light source. In this context, the question that should be investigated in the following sections is whether such an updated version has the ability to outperform Smet's original definition of the memory color rendition index (MCRI) [23, 52–54] considering various lighting conditions.

4.2.6 *Summary (I)*

In the previous sections, a new experiment investigating the influence of long-term memory on the color appearance ratings of familiar objects was presented. By providing realistic viewing and adaptation conditions as well as a more profound test object selection, an attempt was made to overcome the deficiencies observed in previous studies addressing the same topic.

In total, twelve familiar test objects with colors distributed around the hue circle were chosen for the experiments which were performed in a furnished, white-painted experimental room creating a realistic viewing and adaptation scenario. The test objects were individually presented to the observers while their color appearance could be varied over a wide range by applying a combination of a four-channel LED light source and an LCD projector to create the illumination. Special care was taken to ensure a constant adaption state of the observers while they were asked to rate on a five-level preference scale the similarity of the perceived object color with what they thought the respective test object should look like in reality.

For each test object, the pooled observer ratings were subsequently modeled in CIECAM02-UCS chromaticity space using bivariate Gaussian similarity distributions. The centroids of this fitted Gaussian distributions were taken as the corresponding memory color centers which in contrast to the objects' chromaticities calculated under D56 reference illumination tended to be shifted towards higher chroma regions. Exceptions hereto were found for the test objects which are naturally of low saturation such as human complexions and concrete. From the shape and orientation of the similarity distributions it could be further concluded that observers tended to be more tolerant of changes in chroma than of shifts in hue direction.

Comparisons with previous results obtained by Smet *et al.* [40, 53] revealed significant differences in the reported memory color centers but also showed distinct deviations in the covariance matrices defining the size, shape, and orientation of the fitted similarity distribution functions. It is supposed that this new set of similarity functions will lead to significantly different results when being used for the construction of an updated MCRI, which should be investigated and discussed later in this thesis. Furthermore, it is of interest to repeat the experiments at different CCTs of the ambient illumination and with an observer group of different cultural background in order to investigate the influence of both the adapted white point and the cross-cultural differences on memory color assessments. If significant differences between various CCT levels and cultural backgrounds were observed, these effects should be reflected by a universally valid color quality metric being applicable over a wide range of different lighting conditions. The obtained results will be reported in the following.

4.3 IMPACT OF THE ADAPTED WHITE POINT ON MEMORY COLOR ASSESSMENTS

In order to figure out whether or not the adapted white point has a significant impact on the color appearance rating results for the twelve familiar test objects, the whole experiment was repeated at a different CCT of the ambient illumination defining new adaptation conditions.

Table 4.6 – Short- and long-term stability of the new illumination settings for the second round of the experiments where the white point of the ambient illumination is set to 3200 K compared to the original settings at 5600 K. Similar excellent performance can be observed allowing for a proper comparison of the color appearance rating results obtained for the original and the new adaptation conditions.

	Short-term stability in $\Delta u'v'$		Long-term stability in $\Delta u'v'$	
CCT of Ambient Illumination	LCD projector	LED light source	LCD projector	LED light source
3200 K	<0.0019	<0.0010	<0.0037	<0.0017
5600 K	<0.002	<0.0009	<0.004	<0.0015

With a rather cool-white spectrum of the ambient illumination being used in the first run of the experiments, the effect of a more warm-white ambient illumination on the color appearance ratings of the observers should now be considered here. For this purpose, a CCT of 3200 K was chosen since it marks the typical bias point of tungsten halogen fixtures used for theater and film productions [233] as well as for lighting applications in the shopping and retail industry, i.e, in fields of application where the aspirations of providing excellent light sources for high-quality lighting are traditionally strongly pronounced. Further information on the changes made to the original experimental setup will be given in the following. The results of the memory color assessments at the new chromatic adaptation conditions will subsequently be discussed and statistical inference on the impact of the adapted white point on the observers' color appearance ratings will be provided.

4.3.1 *Updated Experimental Conditions*

Basically, the same experimental setup as reported in Sec. 4.2.2 was used for performing the color appearance rating of the familiar objects at 3200 K ambient illumination. Again, three subjects, which were seated on the couch and asked to observe the scene arranged on the table, were tested at the same time. The color appearance of the test objects was changed by the LCD projector, while the ambient illumination was realized by the four-channel LED light source set to a fixed white point as close as possible to the Planckian locus at 3200 K. Regarding the repeatability as well as the short- and long-term stability of this new illumination settings, similar excellent performance as for the original settings was observed. The corresponding upper limits of the measured $\Delta u'v'$ color differences are summarized in Table 4.6. Further details on the photometric properties of the new ambient illumination and the test object selection for the second round of experiments can be found in the following two sections.

4.3.1.1 *Warm-white Ambient Illumination for Chromatic Adaptation*

Like before, the four-channel LED light source setting the ambient light and adaptation conditions was optimized for excellent color rendition given the new CCT of 3200 K. In Fig. 4.20 the resulting normalized spectral radiance of the LED light source is shown in comparison to the 3200 K Planckian radiator which, following the corresponding CIE Technical Report [16, 17, 158], should be taken as the reference illuminant for this specific CCT value. Again, a high general color fidelity index of $R_f = 92.8$ could be achieved with CIE1964 chromaticity coordinates of $x_{10} = 0.4291$ and $y_{10} = 0.3988$ very close to the Planckian locus. The luminances measured on the Spectralon® reflectance target, the white tablecloth, and the white opposing wall were 662.6 cd m^{-2}, 541.5 cd m^{-2}, and 501.7 cd m^{-2}, respectively, which in all three cases corresponds to an increase in luminance of less than 1.8 % when compared to the

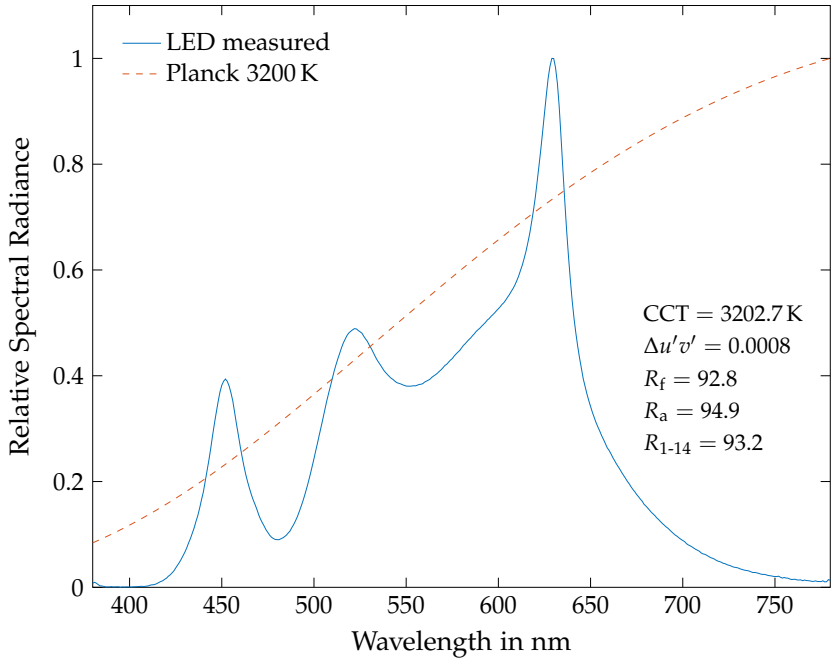


Figure 4.20 – Normalized measured spectral radiance of the optimized four-channel LED light source setting the ambient illumination and adaptation conditions to a CCT of 3200 K. Compared to the Planck 3200 K reference illuminant, excellent color rendition and white point preservation could be achieved.

luminances measured for the cool-white ambient illumination (see Sec. 4.2.2). In addition, the observed illuminances and uniformities caused by the new ambient illumination were also comparable to those obtained for the previous setup showing deviations of less than 1.4 %. Measuring the horizontal illuminance on the coffee table for the new lighting situation for example yielded 2058 lx at a given uniformity of $E_{\text{min}}/\bar{E} = 0.96$, while performing the same measurements on the white opposing wall resulted in a value of 2013 lx with a uniformity of $E_{\text{min}}/\bar{E} = 0.819$.

Due to this excellent agreement of the photometric quantities, an adulterating influence of varying luminance levels of the ambient illumination on the observers' color appearance ratings could be precluded in the present case. Hence, potentially occurring differences in the outcome of both experiments may solely be explained by the different spectral composition of the ambient LED light defining the white point of adaptation.

4.3.1.2 *Similarity of Test Objects and Color Stimuli*

In order to enable a proper comparison between the color appearance rating results obtained for the two different adaptation conditions, special care had to be taken to guarantee similar reflectance properties and visual appearance (when viewed under the same illuminant) of the twelve familiar test objects selected in the second run of the experiments when compared to their originals. For the objects of blue jeans, Asian and Caucasian skin as well as for the concrete flower pot this requirement could easily be fulfilled by simply taking exactly the same object or, in the case of the skin colors, choosing exactly the same hand models for performing the respective experiments. However, for the remaining test objects more effort was necessary. In advance of preparing the experiments for a specific test object, the average spectral reflectance curves of five potential object candidates were (re-)measured as described in Sec. 4.2.3 and compared to the original reflectance curves shown in Fig. 4.14.

Based on these comparisons, the object candidate exhibiting the least deviations regarding its reflectance and appearance characteristics was finally chosen for the subsequent experiments.

For this purpose, the degree of deviation was measured by an adjusted root-mean-square error (RMSE) which is calculated using the following equation:

$$\mathrm{RMSE}_{\mathrm{adj.}} = \sqrt{\frac{1}{N}\sum_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} w(\lambda)\,(\eta_a(\lambda) - \eta_b(\lambda))^2}, \tag{4.19}$$

where $\eta_a(\lambda)$ represents the original spectral reflectance of the test object used during the first run of the experiments, while $\eta_b(\lambda)$ denotes the spectral reflectance of one of the object candidates contemplated for the second run. With the corresponding summation in the present

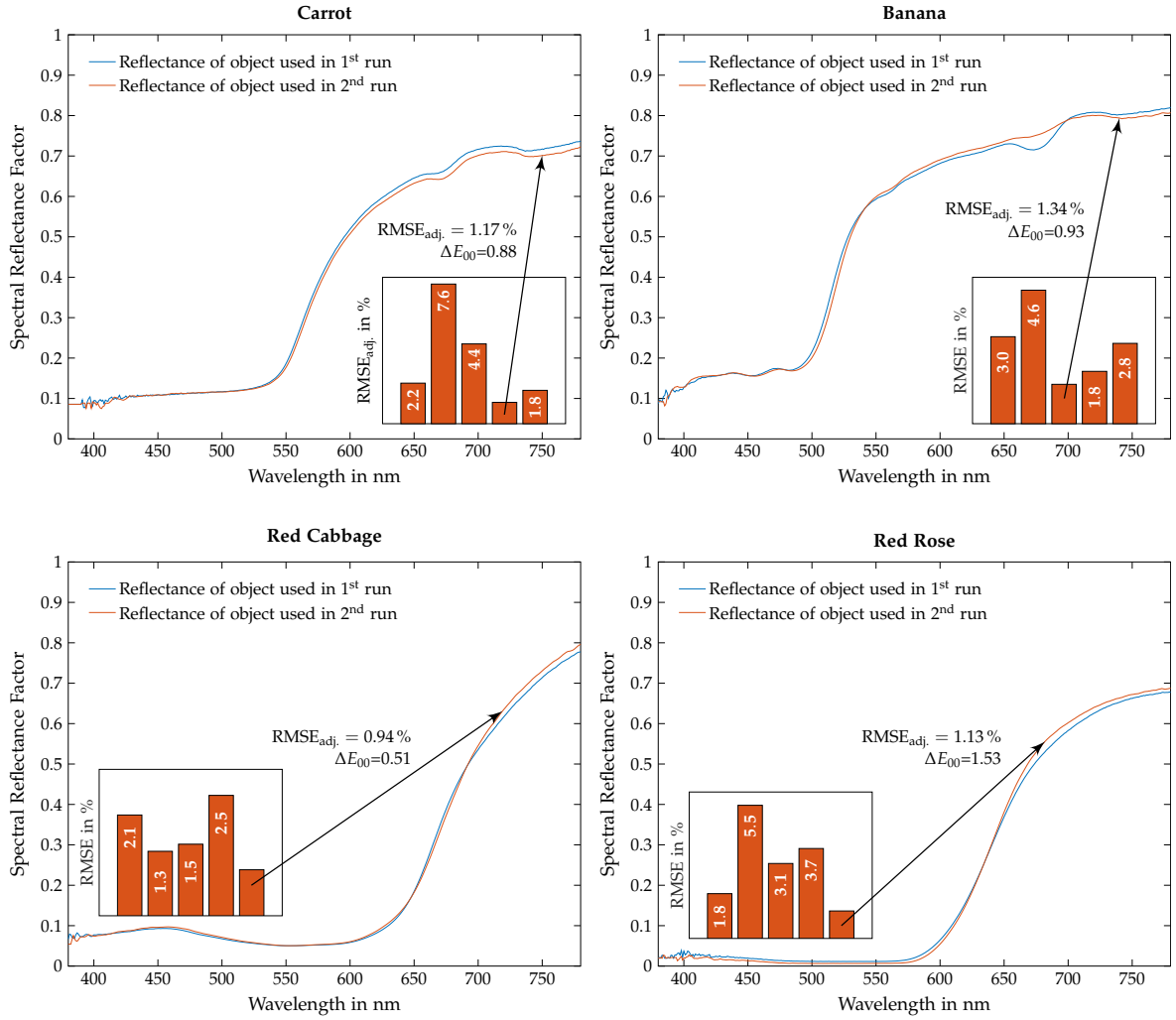


Figure 4.21 – Comparison of the spectral reflectance curve of the original object used in the first run of the experiments (blue line) with the spectral reflectance curve of the object candidate selected for the second run (orange line) for the test objects of carrot (upper left), banana (upper right), red cabbage (lower left), and red rose (lower right). In each case, the additional inset bar chart plot gives the calculated $\mathrm{RMSE}_{\mathrm{adj.}}$ measures of the corresponding five object candidates. From this representation, the one featuring the smallest error was finally chosen for the subsequent experiments which is indicated by the black arrow. In addition, the resulting CIEDE2000 color difference ΔE_{00} as it would be observed when both objects were viewed under the same 3200 K ambient illumination is given for the sake of completeness.

work being performed from 380 nm to 780 nm in steps of 1 nm, the variable N indicating the number of summation steps equaled 401. In addition, a simple filter function of the form

$$w(\lambda) = \begin{cases} 1 & \text{for } 430\,\text{nm} \leq \lambda \leq 780\,\text{nm} \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

had been added to adjust the standard definition of the RMSE in order to avoid the relatively large measurement noise below 430 nm (see Fig. 4.14) to vitiate the validity of the RMSE measure in the current application.

Even though the adjusted RMSE as described by Eq. (4.19) gives a good first order estimate of the similarity of the spectral reflectances, it does not necessarily allow for drawing conclusions about the similarity of the visual appearance of the objects when viewed under

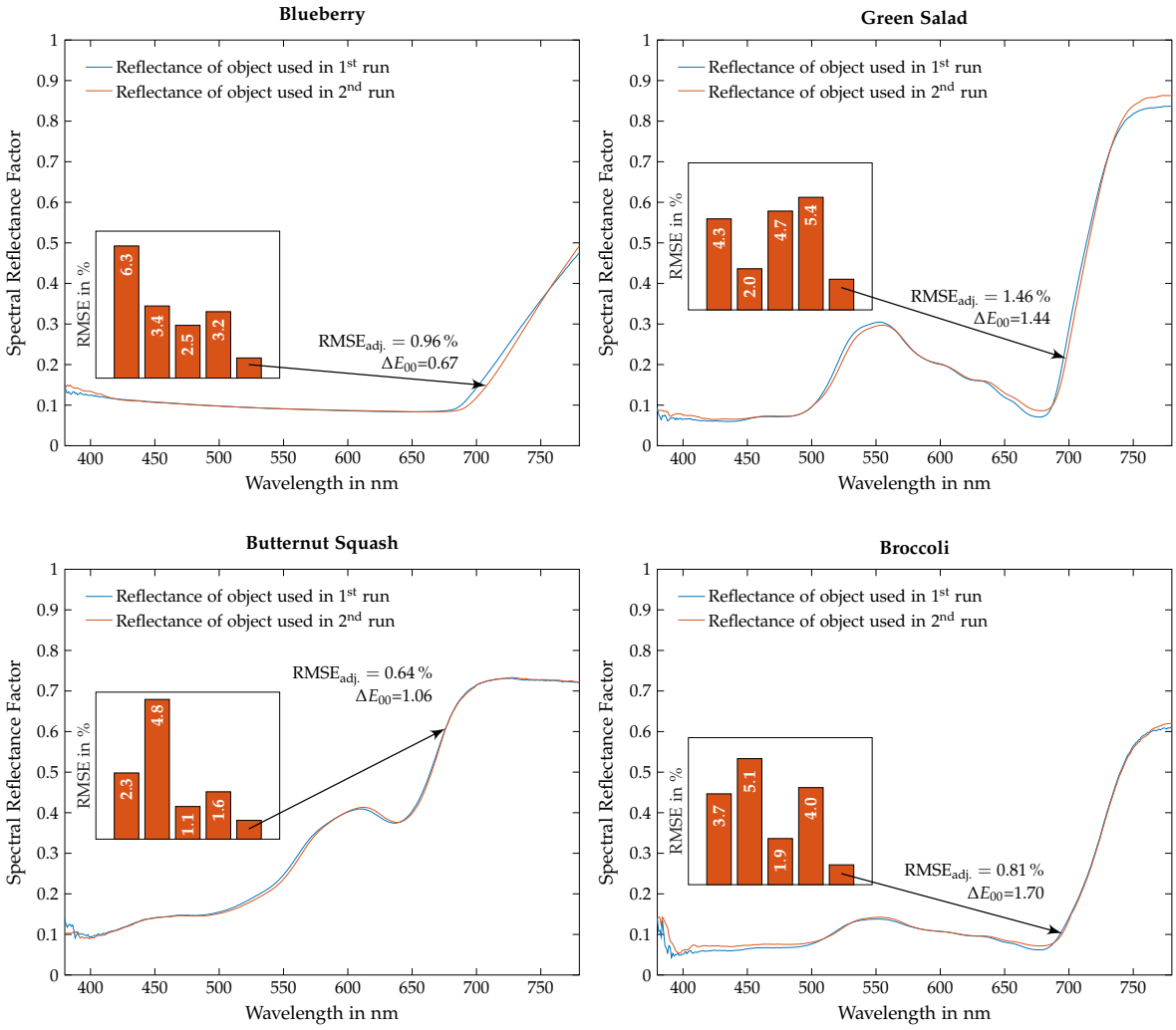


Figure 4.22 – Comparison of the spectral reflectance curve of the original object used in the first run of the experiments (blue line) with the spectral reflectance curve of the object candidate selected for the second run (orange line) for the test objects of blueberry (upper left), green salad (upper right), butternut squash (lower left), and broccoli (lower right). In each case, the additional inset bar chart plot gives the calculated $\text{RMSE}_{\text{adj.}}$ measures of the corresponding five object candidates. From this representation, the one featuring the smallest error was finally chosen for the subsequent experiments which is indicated by the black arrow. In addition, the resulting CIEDE2000 color difference ΔE_{00} as it would be observed when both objects were viewed under the same 3200 K ambient illumination is given for the sake of completeness.

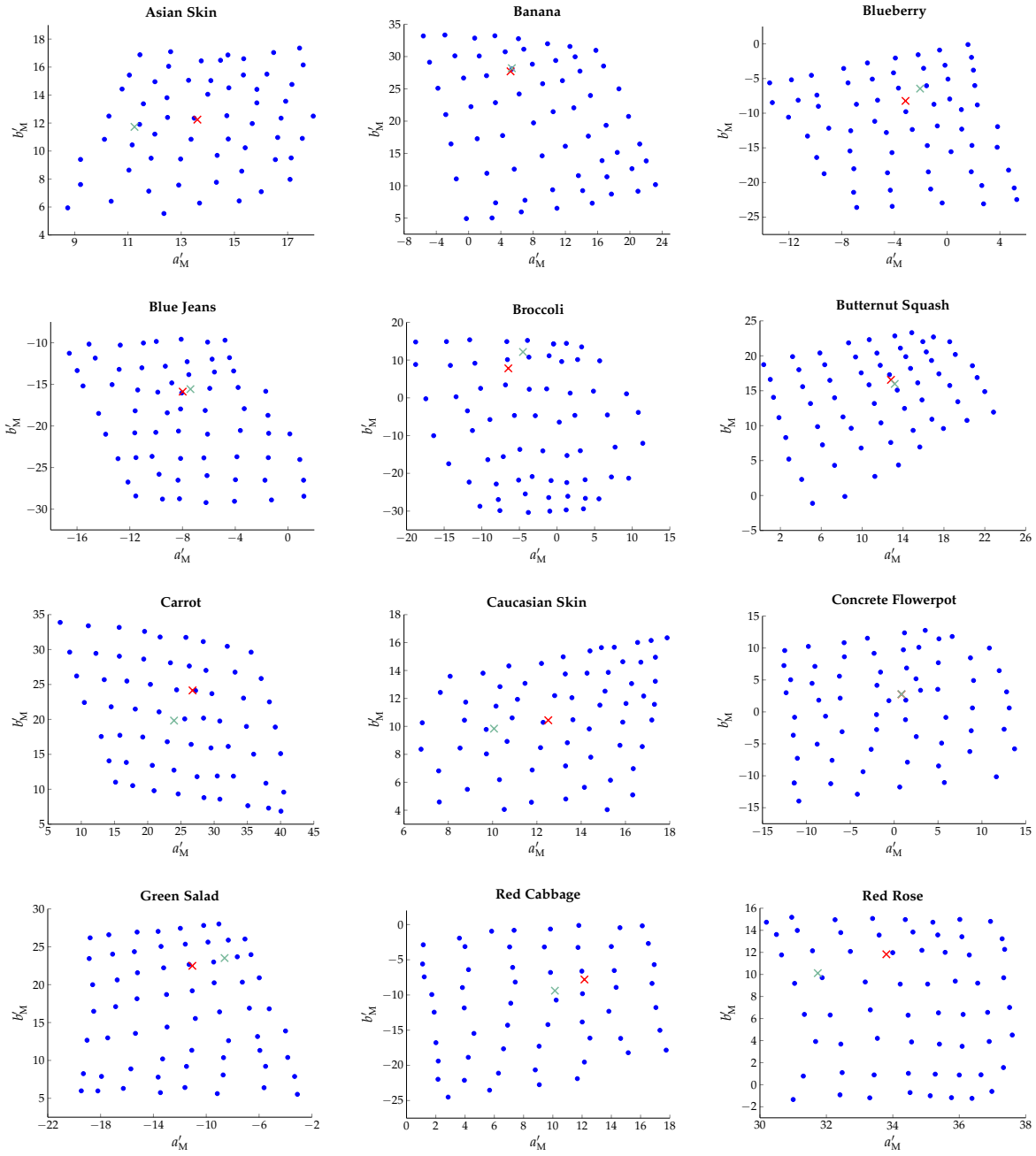


Figure 4.23 – CIECAM02-UCS chromaticities (blue dots) for each of the twelve familiar test objects used for testing the assessment of memory colors at 3200 K ambient illumination. Red crosses mark the objects' chromaticities as they would appear under the 3200 K LED ambient illumination only. For comparative purposes, the light green crosses additionally represent the test objects' chromaticities as they would be perceived under the 5600 K LED ambient illumination adopted in the first run of the experiments (see Sec. 4.2).

identical illumination conditions. Depending on the spectral range in which the major deviations between the two spectral curves occur, one could for example obtain a relatively large RMSE value but a negligibly small perceived color difference or, in the reverse case, one could observe distinct color differences despite of the RMSE measure being quite small. This kind of contradiction is solely due to the underlying nature of the human visual system which, from a colorimetric point of view, is represented by the color matching functions introduced in Sec 2.1. With these functions being most sensitive in the range from 440 nm to 600 nm,

Table 4.7 – Mean values and corresponding $\pm 1\sigma$-intervals of the luminance measure and the CIECAM02-UCS lightness parameter for each test object as observed in the second run of the experiments at 3200 K ambient illumination. The last column tabulates the relative deviations between the mean lightness values of the first and the second run, where positive (negative) numbers indicate an increase (decrease) in lightness.

Test Object	Mean Luminance $\overline{L}$ in cd m^{-2}	Mean Lightness Parameter $\overline{J'}$	$\Delta\overline{J'}$ in %
Asian Skin	302.09 ± 6.09	78.34 ± 0.56	2.85
Banana	503.41 ± 7.68	91.41 ± 0.64	0.42
Blueberry	92.19 ± 2.83	47.56 ± 0.54	1.39
Blue Jeans	76.59 ± 2.35	43.29 ± 0.51	−1.29
Broccoli	104.57 ± 3.00	48.50 ± 0.52	−2.57
Butternut Squash	280.16 ± 6.16	74.34 ± 0.54	1.46
Carrot	287.19 ± 6.26	75.84 ± 0.46	−0.61
Caucasian Skin	376.75 ± 7.91	83.54 ± 0.58	0.54
Concrete Flowerpot	366.32 ± 7.23	81.43 ± 0.53	−2.10
Green Salad	207.38 ± 5.23	64.56 ± 0.53	0.31
Red Cabbage	62.47 ± 1.92	39.88 ± 0.49	−2.23
Red Rose	38.18 ± 0.82	33.09 ± 0.27	2.96

slight deviations in this part of the spectrum may lead to pronounced color differences in the objects' visual appearance, whereas changes in the other parts of the spectrum can be considered as being less severe.

For this reason, colorimetric color differences were additionally calculated in order to get an idea of how much similarity was really observed between the original test objects and the respective object candidates, i.e., not just with regard to their spectral reflectance characteristics but also from a perceptual point of view. Being the current CIE recommendation for computing small color differences [234], which were expected to be observed in the present application, the CIEDE2000 color difference formula as discussed in Sec. 2.2.1 with the reference white of the CIELAB calculation set to the white point of the optimized 3200 K ambient illumination seemed to be a good choice for this purpose.

The final results are summarized in Figs. 4.21 and 4.22 comparing for each test object the spectral reflectance curve used in the original experiments with the spectral reflectance curve of the object candidate that was finally chosen for the second run. As can be seen from the different figures, the absolute and colorimetric deviations between the former and the latter are generally quite small showing $\mathrm{RMSE}_{\mathrm{adj.}}$ values of less than 0.0146 and a maximum ΔE_{00} of only 1.7 for all test objects, which can usually be considered to be negligible. Hence, the postulated similarity of the test objects used in the first and second run of the experiments could be confirmed allowing for a reliable analysis of the impact of the different adaptation conditions at 3200 K and 5600 K on the observers' color appearance ratings.

With the spectral reflectances of the test objects being known, the new object stimuli used for the second run of the experiments were subsequently calculated following the procedure introduced in Sec. 4.2.3. In order to eliminate the presumable influence of different object lightness levels when comparing the observers' chromaticity ratings gathered for different adaptation conditions, the new CIECAM02-UCS chromaticity grids shown in Fig. 4.23 were extracted from the objects' color gamuts in such a way that their mean lightness parameters $\overline{J'}$ matched the original lightness values given in Table 4.1 as good as possible while trying to

keep their corresponding coefficients of variation ($\widehat{=}$ standard deviation of lightness divided by its mean value) smaller than 1 %. As can be seen from Table 4.7, which summarizes the final mean luminance and CIECAM02-UCS lightness values of the twelve test objects as observed in the second run of the experiments, deviations between the mean lightness values of the first and the second run $\Delta\overline{J}' = \left(\overline{J}'_{2nd} - \overline{J}'_{1st}\right) / \overline{J}'_{1st}$ of less than 3 % could be achieved. In the cases of banana, blueberry, blue jeans, carrot, Caucasian skin, and green salad these deviations are even found to be of the same order of magnitude as the corresponding coefficients of variation and, therefore, are considered to be small enough so that any impact of different lightness levels on the observers' color appearance ratings can be precluded.

As described in Sec. 4.2.3, each of the resulting chromaticity grid points shown in Fig. 4.23 corresponds to a certain RGB triplet used for driving the LCD projector which in turn illuminated the respective test object to control its color appearance. Hence, for each test object the accordingly assigned RGB triplets created a test set of stimuli which were presented to the observers in randomized order. Like before, 20 of these stimuli were rated twice within a single test sequence to check for the observers' rating consistency and a small training set of 10 additional RGB triplets was added for the observers to (re-)gain familiarity with the rating scale and the test procedure.

While rating the test objects' color appearance, the chromatic adaptation of the observers was always kept at a constant level, as stated in Sec. 4.2.2, with the adapted white point being defined by the new 3200 K ambient illumination. The outcome of these experiments should be reported in the following section.

4.3.2 *Experimental Results for the New Adaptation Conditions*

For performing the color appearance rating experiments at the new CCT of 3200 K more or less the same panel of subjects could be recruited. Only three of the original participants had to be replaced due to scheduling conflicts. In order to guarantee that the rating data gathered for the two different adaptation conditions were obtained from similar samples representing the same larger, heterogeneous population, they were substituted by specifically chosen surrogates with normal or corrected-to-normal color vision that were of the same age, gender, and origin showing the same level of expertise in color science as their original counterparts.

Before starting the color appearance rating for a specific test object assessed under the new adaptation conditions, observers received exactly the same oral instructions regarding the experimental procedure as discussed in Sec. 4.2.4. Again, they were asked to rate the color appearance of the currently presented familiar test object according to their preference of how they thought the respective object should look like in reality using the identical semi-semantic five-level scale. Once more, they were advised not to stare to long at the test object and asked to look at some scene white each time the emitted projector spectrum was changed in order to ensure the maintenance of a constant chromatic adaptation level during the whole test sequence.

Like in the previous illumination setup, each familiar test object was rated by 15 German observers with an approximately equally balanced male-female ratio resulting in a total number of five test runs being necessary for each test object. With each of these test runs taking just about 20 min to be finished, they could again be efficiently organized and conducted in consecutive order so that the complete testing of a specific familiar object was completed within 2.5 h. As before, no visually perceivable food alteration or degradation was observed during this relatively short period of time.

Table 4.8 – Average inter- and intra-observer PF/3 performance factors and STRESS values calculated from the visual appearance ratings for the twelve familiar test objects at 3200 K adapted white point. In addition, the resulting ICC$(2,k)$ values are shown together with their 95 % confidence intervals. Please note that for both measures PF/3 and STRESS the corresponding values of inter- and intra-observer variability are tabulated in the same column but with the latter given in parenthesis.

Test Object	Inter-(Intra-)observer PF/3	Inter-(Intra-)observer STRESS	ICC$(2,k)$	ICC$(2,k)$ Confidence Interval
Asian Skin	36 (21)	0.29 (0.17)	0.9025	[0.8615, 0.9349]
Banana	32 (17)	0.24 (0.14)	0.9659	[0.9518, 0.9772]
Blueberry	35 (19)	0.28 (0.14)	0.9476	[0.9265, 0.9647]
Blue Jeans	31 (19)	0.25 (0.14)	0.9444	[0.9221, 0.9624]
Broccoli	27 (16)	0.23 (0.14)	0.9780	[0.9692, 0.9850]
Butternut Squash	30 (16)	0.24 (0.14)	0.9685	[0.9554, 0.9789]
Carrot	33 (17)	0.25 (0.14)	0.9613	[0.9451, 0.9742]
Caucasian Skin	38 (22)	0.30 (0.17)	0.9041	[0.8624, 0.9364]
Concrete Flowerpot	31 (16)	0.24 (0.12)	0.9717	[0.9594, 0.9812]
Green Salad	28 (15)	0.22 (0.12)	0.9711	[0.9586, 0.9808]
Red Cabbage	34 (18)	0.26 (0.15)	0.9548	[0.9364, 0.9696]
Red Rose	28 (16)	0.22 (0.12)	0.9608	[0.9447, 0.9737]

The new experiments investigating the assessment of memory colors under a different adaptation condition started right after the original experiments were finished and took another six-weeks period to be completed. Before comparing the results obtained for the different adaptation conditions in the subsequent sections to figure out whether or not significant differences exist, the memory color characteristics that can be derived from the new experiments should be presented first.

Starting with the analysis of the inter- and intra-observer variability to quantify the observers' rating precision and repeatability, Table 4.8 summarizes the PF/3, STRESS, and ICC$(2,k)$ values for the twelve familiar test objects calculated from the visual appearance ratings collected for the new adaptation conditions. As can be seen, the inter- and intra-observer PF/3 measures range from 27 to 38 and from 15 to 22 with an average value of 32 and 18, respectively. Regarding the STRESS measure, values between 0.22 and 0.30 with an average of 0.25 for the inter-observer variability and between 0.12 and 0.17 with an average of 0.14 for the intra-observer variability can be observed, which is in good agreement with the results reported in Sec. 4.2.4.1 for the 5600 K ambient illumination. With the exception of green salad, where both the precision and the repeatability of the observers' color appearance ratings ascertainably increased for the new adaptation conditions, similar inter- and intra-observer measures are obtained for all the remaining test objects when comparing the results of both ambient illumination setups.

Regarding the calculated ICC$(2,k)$ intraclass correlation coefficients, comparably excellent results were achieved in both cases. Like before, ICC$(2,k)$ values larger than 0.90 portend the postulation of an average observer by pooling the individual observer ratings for each test object, which again allows for fitting multivariate Gaussian functions to model the memory color assessments of such an average observer in CIECAM02-UCS chromaticity space. The corresponding results are illustrated in Fig. 4.24 together with the mean observer ratings shown as black dots and the respective goodness-of-fit statistics. Furthermore, the Pearson correlation coefficient ϱ between the model predictions and the mean observer ratings were calculated assuming a 5 % significance level.

As can be stated from the given R^2 values, more than 90 % of the total variations in the mean ratings of the pooled data for each test object can again be explained by the Gaus-

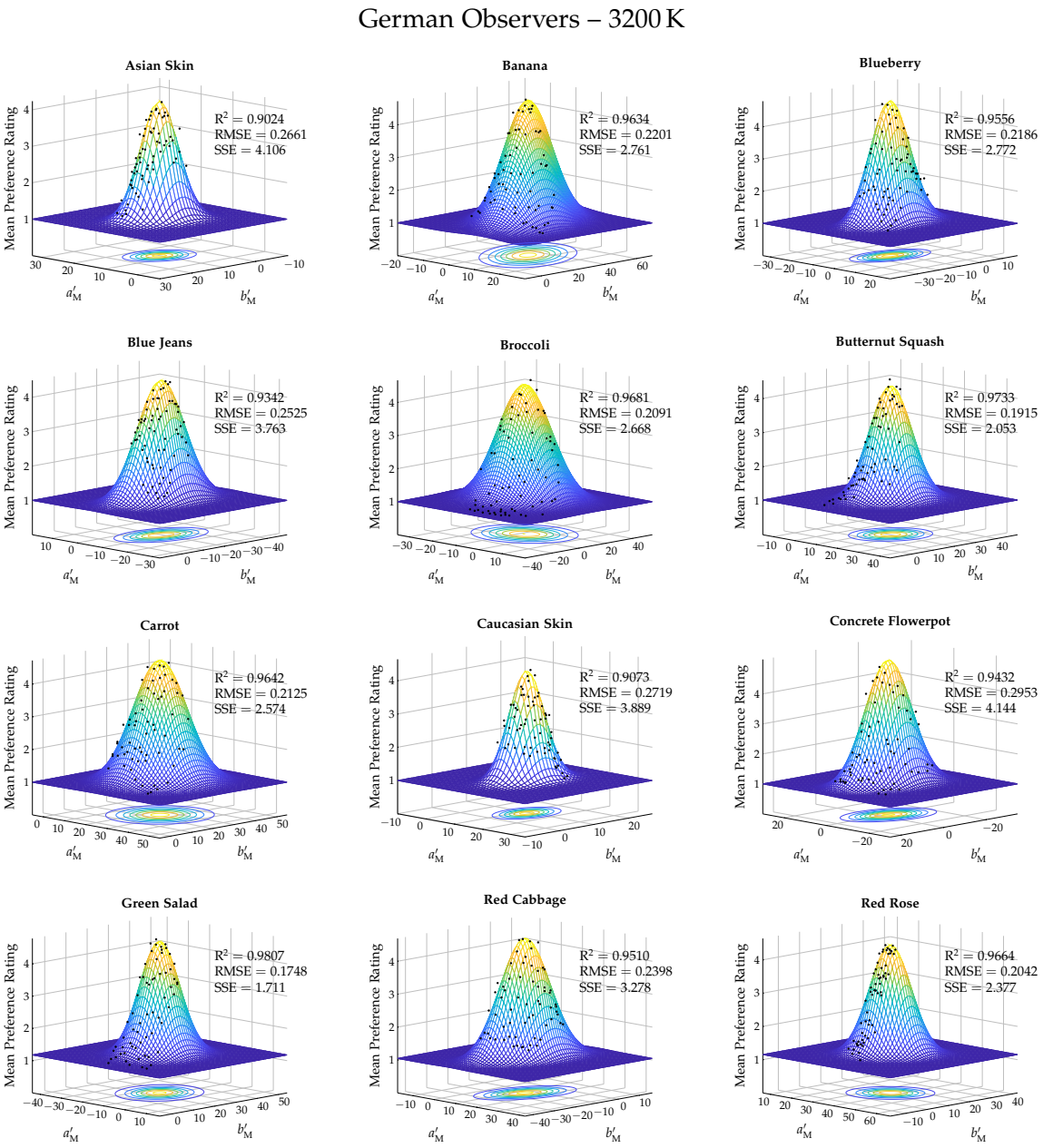


Figure 4.24 – Bivariate Gaussian similarity distributions fitted to the pooled German observer data for each familiar test object modeling the mean preference ratings of an average German observer adapted to the 3200 K ambient illumination in CIECAM02-UCS chromaticity space. Excellent model performance based on the goodness-of-fit statistics can be concluded for all cases. Black dots represent the mean observer ratings for each projector setting whereas the centroids of the Gaussian distributions define the chromaticity coordinates of the corresponding memory colors.

sian modeling, which, in combination with the relatively small RMSE values (< 0.3) and an average Pearson correlation coefficient $\overline{\varrho}$ of 0.97, indicates a comparably excellent model performance as in the case of the 5600 K ambient illumination (see Sec. 4.2.4.2). The chromaticity coordinates of the corresponding memory color centers and the Pearson correlation coefficients are summarized in Table 4.9. Additionally tabulated are the the CIECAM02-UCS chromatic color differences $\Delta E'_{\text{chrom.}}$ between the memory color centers obtained for the two different adaptation conditions. For a better overview, Fig. 4.25 illustrates these color differences in the $(a'_{\text{M}}, b'_{\text{M}})$ chromaticity diagram.

Table 4.9 – CIECAM02-UCS chromaticity coordinates a'_M and b'_M of the memory colors of the twelve familiar test objects given by the centroids of the fitted multivariate Gaussian probability density functions for the 3200 K ambient illumination. Additionally tabulated are the corresponding Pearson correlation coefficients ϱ describing the goodness-of-fit of the Gaussian modeling as well as the CIECAM02-UCS chromatic color differences $\Delta E'_{\text{chrom.}}$ between the memory color centers of both adaptation conditions.

Test Object	a'_M	b'_M	Pearson correlation ϱ	3200 K vs. 5600 K $\Delta E'_{\text{chrom.}}$
Asian Skin	13.04	10.09	0.96	2.76
Banana	5.07	31.77	0.97	0.82
Blueberry	−3.54	−12.62	0.98	0.40
Blue Jeans	−8.11	−19.71	0.97	2.37
Broccoli	−8.62	9.20	0.99	6.19
Butternut Squash	14.73	20.47	0.98	0.66
Carrot	25.09	25.01	0.95	0.83
Caucasian Skin	12.19	9.23	0.98	2.42
Concrete Flowerpot	−0.80	−3.33	0.97	2.33
Green Salad	−13.59	24.19	0.95	4.04
Red Cabbage	12.86	−13.65	0.98	2.34
Red Rose	36.66	12.75	0.98	2.56

By comparing the color centers of each familiar test object assessed by the German observers under 3200 K and 5600 K ambient illumination, deviations between these two adaptation conditions ranging from 0.40 to 6.19 $\Delta E'_{\text{chrom.}}$ can be found. Here, the smallest chromatic color differences are mainly observed for test objects featuring orange to yellowish hues, like for example the test objects of banana, butternut squash, and carrot which all show $\Delta E'_{\text{chrom.}}$ values of less than 0.85. Exception hereto is the object of blueberry which exhibits the smallest color difference of all test objects with $\Delta E'_{\text{chrom.}} = 0.40$. The largest chromatic deviations on

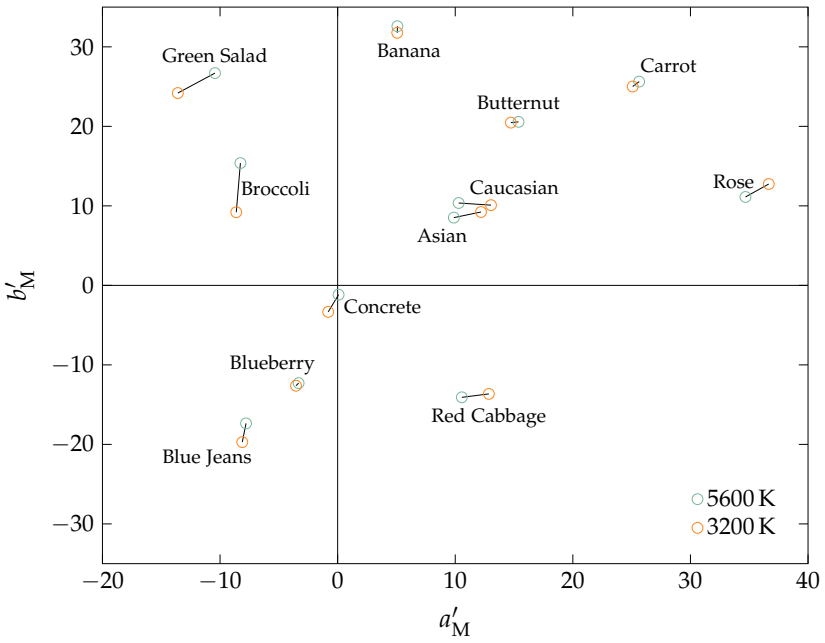


Figure 4.25 – Comparison of the memory color centers of the German observers as obtained by adapting to the 3200 K and 5600 K ambient illumination, respectively. Corresponding memory chromaticities are connected by solid line segments.

the other hand are observed for the test objects in the greenish part of the hue circle, where for example green salad and broccoli give a $\Delta E'_{\text{chrom.}}$ of 4.04 and 6.19, respectively. The overall mean difference $\Delta\overline{E'}_{\text{chrom.}}$ is 2.31, which gives a rough estimate for the order of magnitude of the effect of the adapted white point on the color appearance ratings. With the corresponding color differences being relatively moderate, it can however not yet be confirmed that the different adaptation conditions do have a significant impact. Supplementary analysis is still necessary which will be provided in one of the following sections.

But before getting deeper into the matter of statistics, further characteristics of the test objects' memory colors obtained for the new adaption conditions should be discussed first. For this purpose, Fig. 4.26 compares the chromaticities of the memory color centers reported in Table 4.9 with the hypothetical chromaticities the twelve familiar test objects would show when being solely illuminated by a Planckian reference light source at 3200 K for which the ambient illumination was optimized in the second run of the experiments defining the respective chromatic adaptation conditions. In addition, the corresponding acceptance boundary ellipses are indicated.

When comparing the current results with those illustrated in Fig. 4.18, similar memory-induced chromaticity shifts can be observed. Like in the previous figure, it should be noticed here that most of the chromaticity shifts given by the colored arrows point in the direction of increased CIECAM02-UCS colorfulness M'. Based on the same argumentation as provided in Sec. 4.2.4.3, this memory-induced increase in colorfulness also results in higher CIECAM02 chroma correlates C for most of the test objects. Table 4.10 summarizes these findings in terms of corresponding CIECAM02 chroma increments ΔC, where a plus (minus) sign indicates an

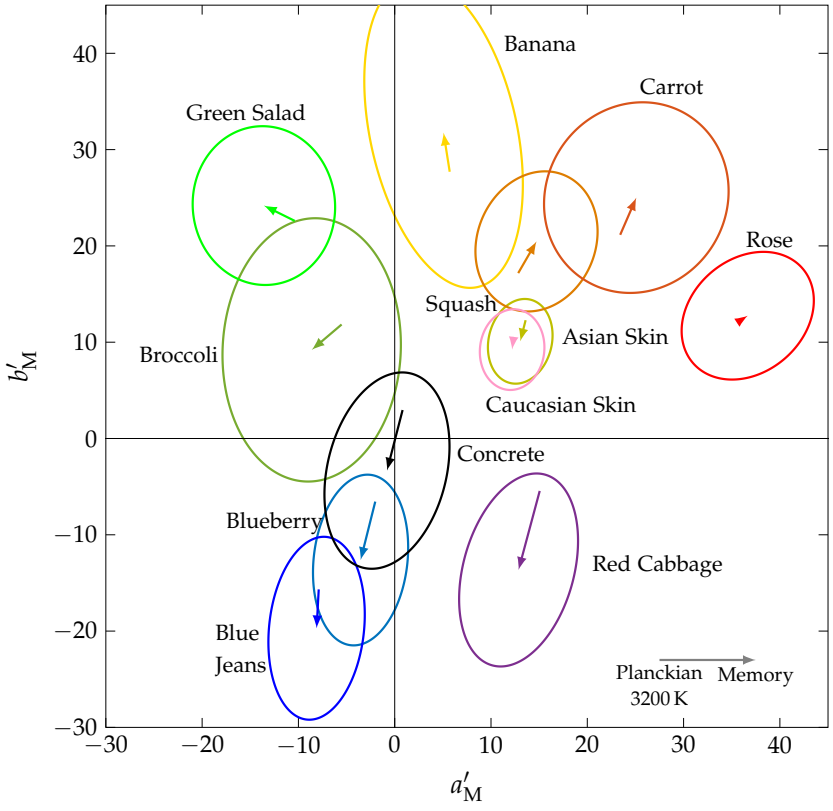


Figure 4.26 – Comparison between the memory color centers and the test object chromaticities rendered using a Planckian reference illuminant at 3200 K. Observed memory-induced chromaticity shifts are indicated by colored arrows. In addition, acceptance boundary ellipses representing an average observer rating of 3 are plotted for each test object.

Table 4.10 – Overview of the chromatic differences and of the deviations in the perception correlates of chroma and hue between the memory color centers and the object chromaticities as perceived under Planckian reference illumination at 3200 K. Additionally tabulated are the geometric measures of the corresponding observer tolerance ellipses. Please note that the semi-major axis a and semi-minor axis b are also given in units of $\Delta E'_{\text{chrom.}}$.

Test Object	$\Delta E'_{\text{chrom.}}$	ΔC	Δh in °	a	b	a/b	θ in °	πab
Asian Skin	2.26	−2.89	−4.37	4.46	3.28	1.36	76.31	45.92
Banana	4.12	8.21	2.58	16.43	7.62	2.16	102.37	393.40
Blueberry	6.25	8.32	1.46	8.90	4.86	1.83	83.36	136.08
Blue Jeans	4.05	6.22	4.36	9.54	4.93	1.94	83.81	147.61
Broccoli	4.08	−0.64	18.17	13.68	9.25	1.48	87.13	397.52
Butternut Squash	3.81	6.83	1.01	7.46	6.13	1.22	67.19	143.67
Carrot	4.20	8.80	2.82	10.10	9.37	1.08	57.72	297.32
Caucasian Skin	1.33	−1.45	−3.36	4.20	3.34	1.26	83.17	44.06
Concrete Flowerpot	6.49	0.41	181.97	10.39	6.18	1.68	76.21	201.67
Green Salad	3.59	5.56	4.66	8.25	7.39	1.12	95.12	191.53
Red Cabbage	8.49	4.32	−26.83	10.28	5.74	1.79	74.52	185.35
Red Rose	1.28	3.15	0.47	7.51	5.88	1.28	40.85	138.88

increase (decrease) in the chroma values of the reported memory color centers in comparison to the chroma values the test objects would show under reference illumination.

As can be seen, the general tendency of memory colors to be shifted towards higher chroma can again be confirmed with an average $\Delta\overline{C}$ of ~4. Furthermore, for most of the test objects these increments are of the same order of magnitude regarding the two different adaptation conditions. Thus, it can be concluded that the observed cognitive effect of claiming an additional amount of saturation in the objects' color appearance to match the observers' expectations is to some extent independent of the adapted white point, at least for the phosphor-converted LED spectra used in the present study (when using narrow-banded LEDs for color mixing to achieve a certain white point things might be a little bit different due to the missing light emission in some parts of the mixed spectrum).

Similar conclusions can be drawn for the hue shifts Δh and chromatic color differences $\Delta E'_{\text{chrom.}}$ calculated between the familiar test objects' reference representations and their respective memory color centers, which are also tabulated in Table 4.10. For most of the test objects assessed under the 3200 K ambient illumination, the corresponding memory colors are again shifted towards the typical hues commonly associated with the respective objects. Besides being oriented in the same direction, these hue shifts in most cases are approximately of the same order of magnitude as obtained for the adaptation conditions at 5600 K, which can be seen by comparing Tables 4.10 and 4.4, respectively. Regarding the chromatic color differences $\Delta E'_{\text{chrom.}}$, which, as stated in Sec. 4.2.4.3, provide a basic measure for the strength of the long-term memory effects on the color appearance ratings, an average value of 4.16 $\Delta E'_{\text{chrom.}}$ must be reported for the 3200 K ambient illumination. This average chromatic difference is in pretty good agreement with the one found for the 5600 K setup showing a value of 4.19 $\Delta E'_{\text{chrom.}}$.

Even though comparable tendencies of the memory-induced chromaticity shifts can be concluded for the two different adaptation conditions, it should be emphasized that these conclusions do not hold true for all familiar test objects. In particular, the test object of

broccoli shows somewhat larger deviations. While in case of the 5600 K ambient illumination its memory-induced chromaticity shift points strongly into the direction of increased chroma, a rather hue-oriented shift must be reported for the 3200 K adaptation condition giving the broccoli a slightly more bluish color appearance than it would actually be perceived under the Planckian radiator. Nevertheless, the general trend that memory colors of familiar objects in comparison to their appearance under reference illumination tend to be shifted towards higher chroma as well as towards their typical hues could still be confirmed and seems to be at least to some degree independent of the observers' chromatic state of adaptation. Please note that, in contrast to this general tendency observed for almost all memory colors, the test objects of Asian and Caucasian skin, that are naturally of lower saturation, exhibit again a decrease in chroma as well as a shift towards a slightly redder hue, which is obviously in accordance to the findings obtained for the 5600 K ambient illumination and should therefore be considered as a more universal feature of these kind of memory colors rather than being an experimental artifact.

In addition to the descriptive analysis of the impact of the adapted white point on the general characteristics of the memory-induced chromaticity shifts provided above, it is also of interest to compare the acceptance boundary ellipses derived from the respective fitted similarity distributions. Even though some slight differences in terms of ellipse location, size, shape, and orientation can be reported from the comparison of the corresponding geometric ellipse measures given in Tables 4.5 and 4.10 or from a closer inspection of Figs. 4.18 and 4.26, respectively, the general pattern of the acceptance boundary ellipses remains more or less the same for the two different ambient illuminations. Hence, with the observed deviations between the varying adaptation conditions being less pronounced than it might be expected at first glance, the question still remains whether or not the adapted white point has a systematic and significant impact on the assessment of memory colors. In the attempt of providing an answer, a more sophisticated statistical analysis of the color appearance rating data is required and should be performed in the two subsequent sections.

4.3.3 *Statistical Inference on Observer Variability*

In order to be able to identify a potential impact of the adapted white point on the observer variability, the distributions of the inter- and intra-observer variabilities in terms of PF/3 and STRESS units for the two different adaptation conditions should be compared and tested for significance. In order to get an idea about the nature of the underlying population of the present data samples, Fig. 4.27 compares the distributions of the individual inter- and intra-observer PF/3 values calculated for each test object as obtained for the two different adaptation conditions at 3200 K and 5600 K, respectively, using the standard box plot data visualization approach. In Fig. 4.28 the same comparison is provided but only for the STRESS instead of the PF/3 measure.

As can be seen from both figures, no real systematic trends can be observed between the two adaptation conditions. Both the interquartile range (IQR) and the whole data range of the underlying distributions given by the ends of the respective whiskers vary quite strongly from test object to test object. In some cases, the observer variability data obtained for the 3200 K ambient illumination shows a smaller IQR range and, therefore, less variance than their 5600 K counterparts. In other cases, the opposite holds true. Similar observations can be reported for the median of the distributions. Hence, even if significant differences between the variability distributions of the test objects for the two different adaptation conditions could be found, the data still lack of systematic tendencies of for example a general decrease

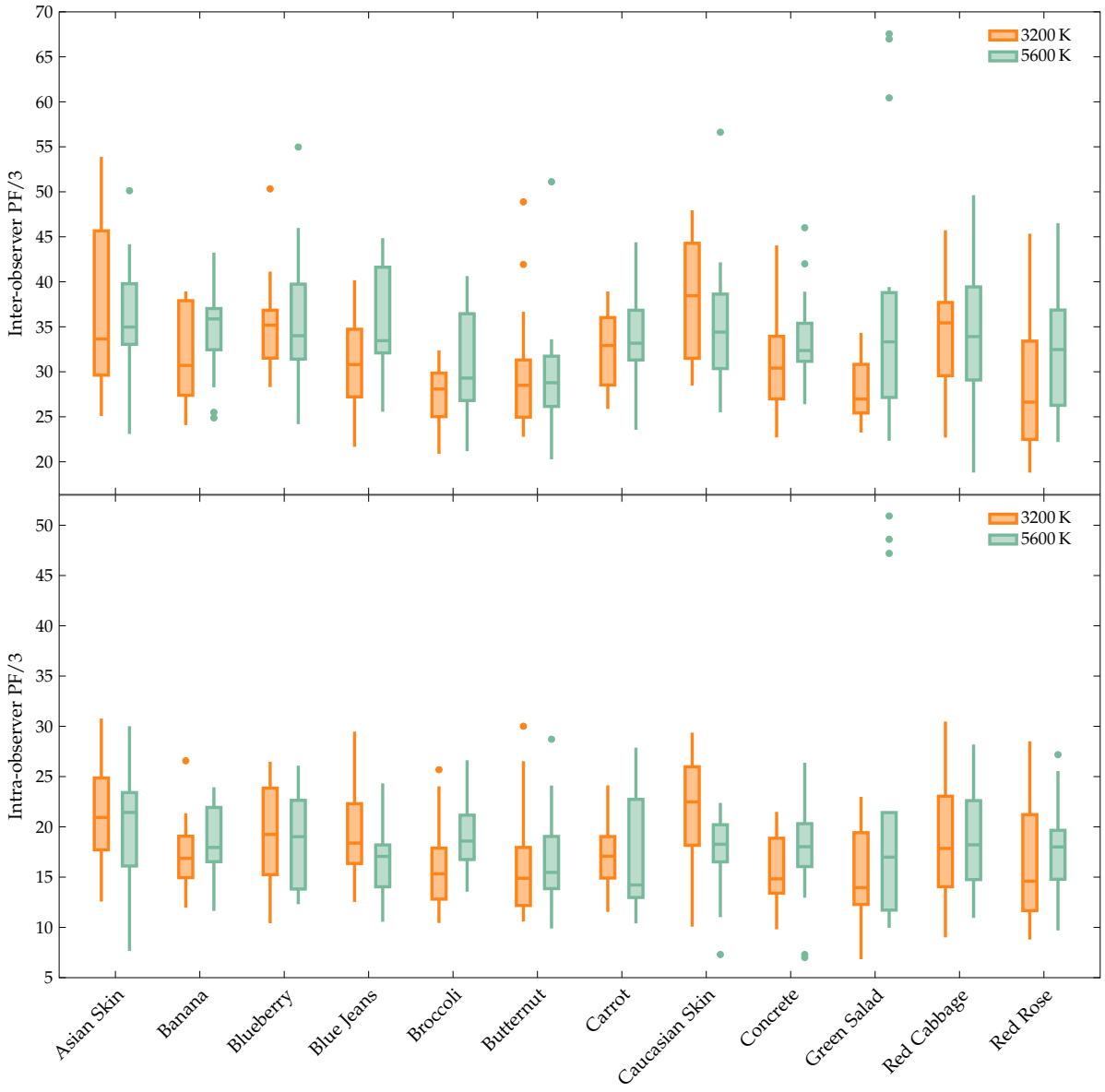


Figure 4.27 – Box plot comparison of the distributions of the individual inter-observer (upper plot) and intra-observer (lower plot) PF/3 values calculated for each combination of test object and adaptation condition. Here, the line inside the box gives the median of the distribution, while the bottom and top of the box represent its 25 % and 75 % quartile. The end of the whiskers indicate the lowest and highest data point lying at most 1.5 times the interquartile range below the lower and above the upper quartile, respectively. Any data point not included between the whiskers is identified as a suspected outlier and indicated by a colored dot.

in observer variability in the color appearance ratings when being adapted to a certain illumination condition.

However, what should be noticed is a quite good correlation between the individual PF/3 and STRESS measures regarding both the inter- and intra-observer variability. In most cases, observing a small (large) IQR for the inter-observer PF/3 distribution of a certain test object basically gives a small (large) IQR for the respective inter-observer STRESS distribution, while approximately retaining the same ratio $\mathrm{IQR}_{3200K}/\mathrm{IQR}_{5600K}$ between both adaptation conditions. A similar behavior is found for the intra-observer variability measures. For a better visualization, Fig. 4.29 illustrates the good to excellent correlation (Pearson $\varrho > 0.90$)

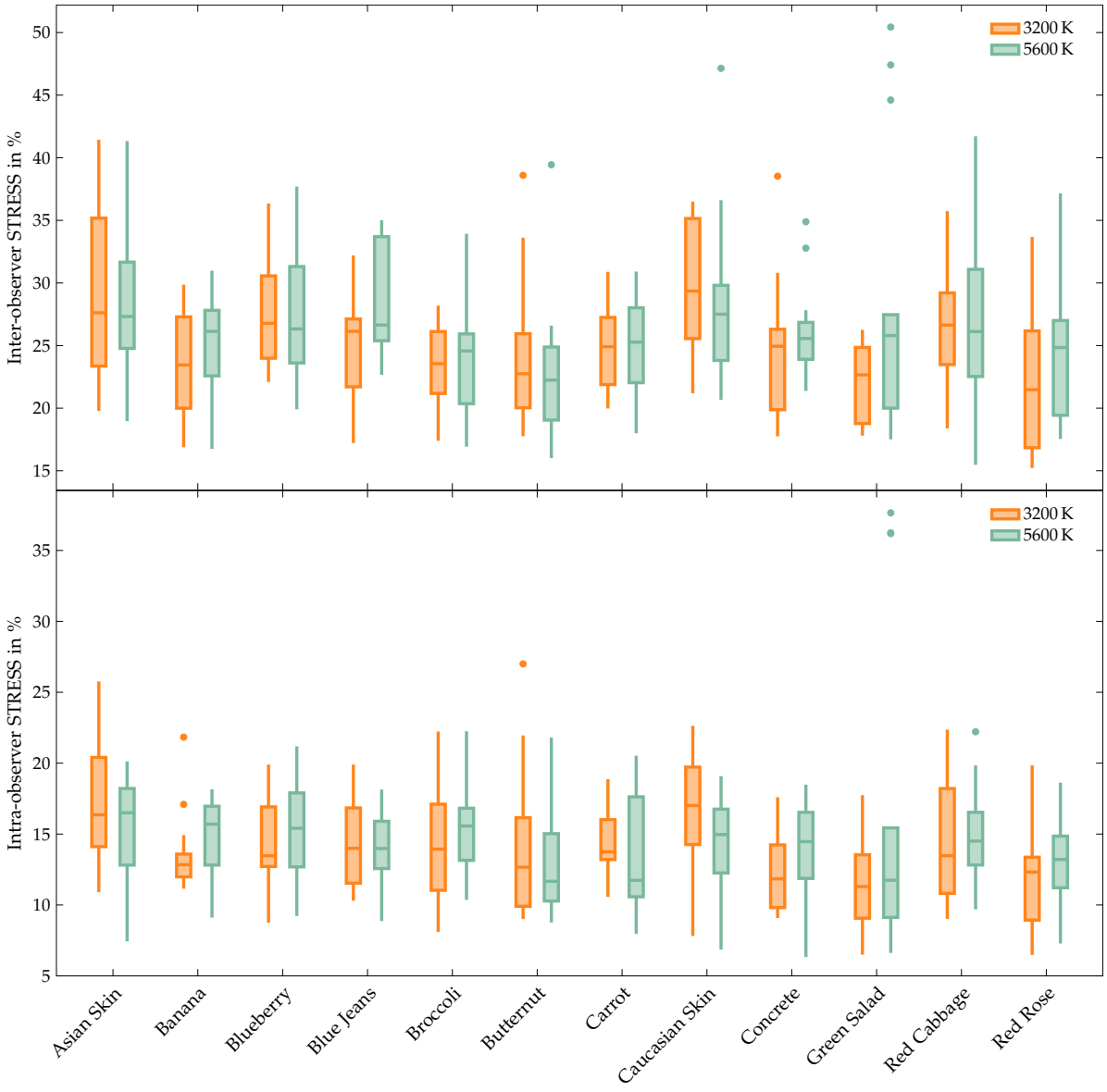


Figure 4.28 – Box plot comparison of the distributions of the individual inter-observer (upper plot) and intra-observer (lower plot) STRESS values calculated for each combination of test object and adaptation condition. Here, the line inside the box gives the median of the distribution, while the bottom and top of the box represent its 25 % and 75 % quartile. The end of the whiskers indicate the lowest and highest data point lying at most 1.5 times the interquartile range below the lower and above the upper quartile, respectively. Any data point not included between the whiskers is identified as a suspected outlier and indicated by a colored dot.

between the pooled STRESS and PF/3 measures for all possible combinations of observer variability type and adapted white point. On the other hand, no correlation was found between the inter- and intra-observer PF/3 values or between the inter- and intra-observer STRESS values for the same ambient illumination, i.e., a small (large) PF/3 or STRESS measure of one kind of variability type does not necessarily lead to a small (large) PF/3 or STRESS measure of the other kind of variability type.

Regarding the further analysis of the variability distributions visualized in Figs. 4.27 and 4.28, it is of interest to first have a look at the data points lying outside the range of the box plot whiskers that can be identified for some of the test objects. These so-called suspected outliers are further investigated using the generalized extreme Studentized deviate (ESD)

test introduced by Rosner [236, 237], which in contrast to other alternatives only requires an upper bound for the suspected number of outliers to be specified. Even though it is essentially a sequential application of the Grubbs test [238, 239], the generalized ESD test is considered to be more reliable and accurate since it makes an appropriate adjustment of the critical values based on the number of outliers being tested which the sequential application of the Grubbs test would not provide. By running the generalized ESD test, clear outliers could be identified and removed from the data sets. Next, the corrected variability distributions were tested for normality.

For this purpose, Kolmogorov-Smirnov (KS) [235, 240–242] and Shapiro-Wilk (SW) [241, 243, 244] tests were applied to the present data. While the former test is pretty common in various fields of statistical analysis, the latter offers superior test power even for a small sample size [245, 246], which obviously is the case here. The results of these normality tests applied to the observer variability distributions obtained for the different test objects are shown in Tables 4.11 and 4.12 for the PF/3 and in Tables 4.13 and 4.14 for the STRESS measure together with the respective z-scores of skewness and excess kurtosis [247–249].

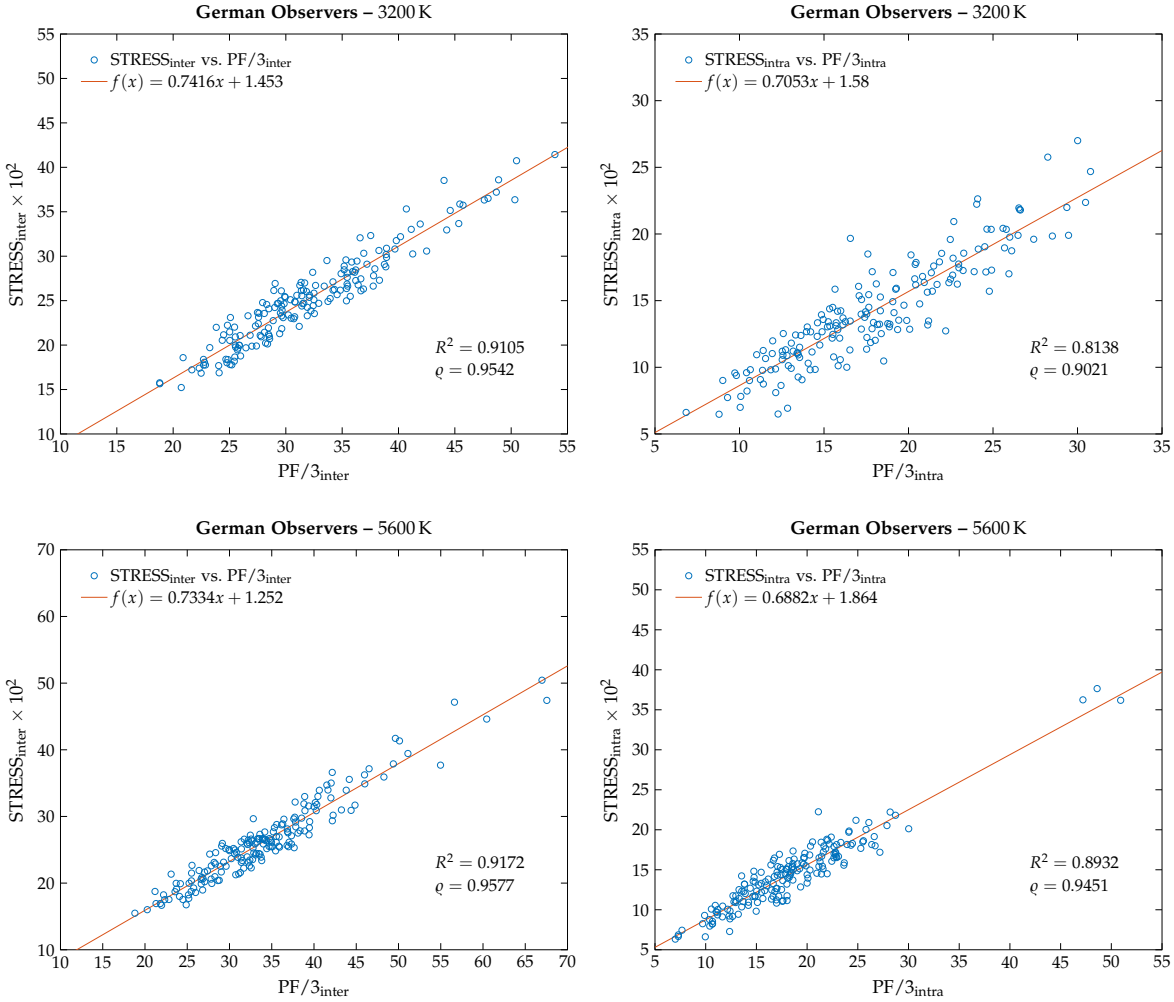


Figure 4.29 – Scatter plots for each combination of variability type (inter vs. intra) and adapted white point (3200 K vs. 5600 K) illustrating the good to excellent positive correlation between the pooled STRESS and PF/3 measures. In all cases, a Pearson correlation coefficient of $\varrho > 0.90$ can be reported, while the general trend of the data is well described by the respective line of best fit giving R^2 values ranging between 0.81 and 0.92.

Table 4.11 – Results of various normality tests applied to the data distributions of the inter- and intra-observer PF/3 measure for German observers obtained at 3200 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Inter-observer PF/3 at 3200 K						Intra-observer PF/3 at 3200 K					
			KS[1] test		SW test				KS[1] test		SW test	
	z_{skewness}	z_{kurtosis}	statistic	p value	statistic	p value	z_{skewness}	z_{kurtosis}	statistic	p value	statistic	p value
Asian Skin	1.081	−0.896	0.202	0.096	0.890	0.068	0.488	−0.285	0.127	>0.5	0.968	0.826
Banana	0.180	−1.336	0.157	0.441	0.895	0.097	0.086	−0.582	0.076	>0.5	0.993	1.000
Blueberry	0.439	−0.621	0.175	0.283	0.964	0.792	−0.112	−0.988	0.126	>0.5	0.955	0.598
Blue Jeans	−0.056	−0.471	0.103	>0.5	0.986	0.995	1.031	−0.064	0.157	0.389	0.967	0.814
Broccoli	−0.667	−0.831	0.145	>0.5	0.955	0.603	1.482	0.258	0.153	0.488	0.933	0.283
Butternut	1.066	0.122	0.115	>0.5	0.943	0.414	2.155	1.224	0.161	0.396	0.877	0.051
Carrot	−0.202	−1.093	0.141	>0.5	0.943	0.423	0.577	−0.597	0.145	>0.5	0.961	0.705
Caucasian	−0.139	−1.197	0.154	0.472	0.934	0.348	−1.212	−0.058	0.135	>0.5	0.951	0.545
Concrete	1.089	−0.255	0.114	>0.5	0.957	0.673	0.278	−1.009	0.157	0.433	0.950	0.567
Green Salad	0.599	−1.007	0.185	0.215	0.931	0.314	0.432	−0.598	0.179	0.252	0.947	0.514
Red Cabbage	−0.140	−0.293	0.203	0.093	0.956	0.618	0.361	−0.607	0.134	>0.5	0.973	0.896
Red Rose	1.174	−0.172	0.155	0.457	0.939	0.410	1.111	−0.548	0.156	0.393	0.929	0.267

[1] with Lilliefors correction [235]

Table 4.12 – Results of various normality tests applied to the data distributions of the inter- and intra-observer PF/3 measure for German observers obtained at 5600 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Inter-observer PF/3 at 5600 K						Intra-observer PF/3 at 5600 K					
			KS[1] test		SW test				KS[1] test		SW test	
	z_{skewness}	z_{kurtosis}	statistic	p value	statistic	p value	z_{skewness}	z_{kurtosis}	statistic	p value	statistic	p value
Asian Skin	−0.492	−0.339	0.186	0.310	0.960	0.730	−0.837	−0.064	0.186	0.170	0.966	0.803
Banana	−0.295	0.695	0.146	0.243	0.908	0.147	−0.160	−0.793	0.146	>0.5	0.955	0.606
Blueberry	0.313	−0.364	0.217	>0.5	0.964	0.781	−0.011	−1.382	0.217	0.053	0.886	0.059
Blue Jeans	0.178	−1.038	0.177	0.144	0.926	0.237	0.554	−0.529	0.177	0.225	0.950	0.530
Broccoli	0.456	−0.921	0.129	0.228	0.920	0.192	0.836	−0.288	0.129	>0.5	0.973	0.905
Butternut	−0.794	−0.791	0.152	>0.5	0.934	0.352	0.732	−0.327	0.152	0.498	0.962	0.750
Carrot	−0.083	−0.131	0.231	0.335	0.955	0.612	0.935	−1.071	0.231	0.030	0.883	0.053
Caucasian	−0.302	−0.943	0.171	>0.5	0.965	0.802	−1.675	0.241	0.171	0.314	0.883	0.061
Concrete	0.129	0.147	0.121	>0.5	0.966	0.761	0.748	−0.087	0.121	>0.5	0.971	0.904
Green Salad	0.059	−0.959	0.205	>0.5	0.968	0.890	0.006	−1.079	0.205	0.171	0.906	0.189
Red Cabbage	0.276	−0.696	0.124	>0.5	0.957	0.643	0.151	−0.773	0.124	>0.5	0.964	0.754
Red Rose	0.805	−0.372	0.122	0.246	0.951	0.546	0.484	0.055	0.122	>0.5	0.973	0.842

[1] with Lilliefors correction [235]

Basically, skewness and excess kurtosis are measures of shape. Whereas the former gives an idea about the amount and direction of skew in the data samples relative to a normal distribution, the latter provides an estimate of how tall and sharp the central peaks of the respective data distributions are. However, they can both be used to test for normality. With a normal distribution having skewness and excess kurtosis of zero, it is very likely for a sample distribution to be drawn from a normal distributed population if its sample skewness and excess kurtosis are also close to zero. On the other hand, the assumption of normality must be rejected when the sample distribution is too much skewed or deviates too much from normal kurtosis to be explained by random chance. By dividing the sample skewness and excess kurtosis by their corresponding standard errors, a test statistic based on z-scores can

Table 4.13 – Results of various normality tests applied to the data distributions of the inter- and intra-observer STRESS measure for German observers obtained at 3200 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Inter-observer STRESS at 3200 K						Intra-observer STRESS at 3200 K					
			KS[1] test		SW test				KS[1] test		SW test	
	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value
Asian Skin	0.942	−0.896	0.166	0.307	0.912	0.147	0.884	−0.556	0.132	>0.5	0.951	0.545
Banana	0.014	−0.971	0.132	>0.5	0.957	0.666	0.936	−0.128	0.129	>0.5	0.937	0.456
Blueberry	0.915	−0.660	0.163	0.337	0.943	0.424	−0.082	−0.819	0.134	>0.5	0.966	0.799
Blue Jeans	−0.276	−0.272	0.169	0.286	0.974	0.915	0.281	−1.021	0.129	>0.5	0.946	0.459
Broccoli	−0.507	−0.886	0.127	>0.5	0.957	0.644	0.587	−0.647	0.128	>0.5	0.963	0.748
Butternut	1.632	0.719	0.113	>0.5	0.914	0.158	1.620	0.004	0.175	0.282	0.890	0.075
Carrot	0.061	−0.794	0.153	0.425	0.939	0.372	0.658	−0.584	0.182	0.197	0.941	0.390
Caucasian	−0.103	−1.004	0.152	>0.5	0.944	0.467	−0.919	−0.365	0.152	0.433	0.952	0.558
Concrete	0.179	−0.782	0.178	0.301	0.931	0.355	1.051	−0.625	0.146	>0.5	0.916	0.193
Green Salad	−0.161	−1.400	0.200	0.128	0.883	0.065	0.298	−0.866	0.108	>0.5	0.944	0.465
Red Cabbage	0.202	−0.307	0.118	>0.5	0.983	0.985	0.526	−1.083	0.145	>0.5	0.935	0.326
Red Rose	0.828	−0.683	0.193	0.166	0.924	0.250	0.704	−0.550	0.175	0.245	0.944	0.436

[1] with Lilliefors correction [235]

Table 4.14 – Results of various normality tests applied to the data distributions of the inter- and intra-observer STRESS measure for German observers obtained at 5600 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Inter-observer STRESS at 5600 K						Intra-observer STRESS at 5600 K					
			KS[1] test		SW test				KS[1] test		SW test	
	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value
Asian Skin	1.075	0.227	0.133	>0.5	0.958	0.570	−1.236	−0.163	0.156	0.394	0.926	0.238
Banana	−1.204	−0.158	0.193	0.135	0.949	0.502	−1.369	−0.334	0.158	0.380	0.916	0.167
Blueberry	0.674	−0.699	0.161	0.348	0.952	0.559	−0.166	−0.944	0.162	0.342	0.951	0.544
Blue Jeans	0.266	−1.364	0.207	0.079	0.884	0.055	−0.481	−0.602	0.160	0.358	0.944	0.430
Broccoli	0.764	−0.552	0.173	0.259	0.950	0.521	0.679	−0.104	0.121	>0.5	0.971	0.874
Butternut	−0.419	−1.052	0.139	>0.5	0.940	0.416	1.626	0.006	0.189	0.152	0.902	0.094
Carrot	−0.512	−0.774	0.095	>0.5	0.968	0.821	0.780	−1.048	0.204	0.089	0.908	0.125
Caucasian	1.014	−0.155	0.144	>0.5	0.956	0.659	−1.218	−0.246	0.196	0.122	0.942	0.411
Concrete	−0.679	−0.810	0.154	>0.5	0.946	0.535	−1.340	−0.254	0.122	>0.5	0.916	0.169
Green Salad	−0.256	−1.349	0.233	0.070	0.858	0.046	0.134	−0.725	0.151	>0.5	0.944	0.548
Red Cabbage	0.772	−0.467	0.127	>0.5	0.970	0.864	0.126	−0.277	0.132	>0.5	0.970	0.873
Red Rose	1.323	0.277	0.144	>0.5	0.915	0.140	−0.213	−0.301	0.110	>0.5	0.982	0.983

[1] with Lilliefors correction [235]

be derived [249]. It can be shown that with the assumption of a 5 % significance level the null hypothesis of normal distributed data must be rejected if the absolute z-scores of either sample skewness or excess kurtosis are larger than 1.96 [247].

From Tables 4.11 to 4.14 it can be seen that the assumption of dealing with normally distributed inter- and intra-observer variability data holds for most of the familiar test objects assessed under both adaptation conditions, which also justifies the previous application of the generalized ESD test to correct for outliers in the data. However, slight inconsistencies are observed for the objects of butternut squash, carrot, and green salad. By evaluating for example the individual color appearance ratings of the test object of butternut squash, it should be noticed that the variability distribution of the individual intra-observer PF/3 measures

at 3200 K adapted white point shows to much skewness for being normal. For the object of carrot, on the other hand, it is less about the shape of the respective variability distribution than about the conclusions drawn from the outcome of the statistical inspection: With the p value of the applied KS test being less than 0.05, indication is given that the corresponding intra-observer PF/3 data at 5600 K ambient illumination are not normally distributed. Here, a 5 % significance level is assumed. Last but not least, a similar result is obtained for the test object of green salad regarding the inter-observer variability distribution of the individual STRESS measures at 5600 K, where the SW test gives a p value smaller than 0.5 again leading to the rejection of the null hypothesis of normally distributed data at a 5 % significance level. Since the assumption of normality is partly violated for these three familiar test objects, both parametric and non-parametric statistical tests should be applied in the following in order to determine if there is a significant difference in the observer variability of the present color appearance rating experiments caused by the different adaptation conditions.

Due to its simplicity and robustness even to slight departures from the assumption of normality [250–253], a two-sample t-test [254–256] is taken for parametric testing, while the Wilcoxon rank-sum test [257–259] is chosen as the non-parametric alternative, which unlike the t-test does not require normally distributed data but is nearly as efficient when applied to normal distributions [260]. However, the application of the standard versions of these tests to the present data also demands equal variances (homoscedasticity) of the variability distributions that should be compared with each other. In this context, a reliable and common method to test for homoscedasticity providing good robustness against many types of non-normal data while retaining excellent test power is the Brown-Forsythe extension of Levene's test [261] which uses an approximated F-statistic to test the null hypothesis that the variances are equal across the distributions under inspection [262, 263].

By applying the Brown-Forsythe extension of Levene's (BFL) test to the present data sets, the null hypothesis of equal variances must be rejected at a 5 % significance level for the inter-observer PF/3 variability distributions of banana ($F = 6.50$, $p < 0.018$), for the inter-observer STRESS variability distributions of concrete ($F = 4.67$,$p < 0.041$), and for the intra-observer STRESS variability distributions of banana ($F = 4.43$, $p < 0.046$) and red cabbage ($F = 4.21$, $p < 0.0499$). For the remaining test objects and data types, the assumption of equal variances between the respective variability distributions obtained for the two different adapted white points can be confirmed.

Hence, in case of normally distributed data where homoscedasticity is met the standard definition of the two-sample t-test is applied to test for significant differences in the observer variability potentially caused by the different adaptation conditions. If the assumption of equal variances is violated (Behrens-Fisher Problem, [264–267]) but the data are still normally distributed, the so-called Welch-test [268, 269], a modified version of the t-test for unequal variances, is chosen as the alternative. For non-normally distributed data, the Wilcoxon rank-sum test can be used natively, i.e., no heteroscedastic version is necessary for the present data sets. Table 4.15 summarizes the findings for all test objects assessed under both adaptation conditions.

As can be seen, significant differences in the inter-observer variability data between the two different adaptation conditions are only observed for the test objects of banana in terms of the PF/3 measure and for the test object of blue jeans regarding both the STRESS and the PF/3 measure. For the remaining test objects, the null hypothesis that the sampled variability data for the two different adapted white points have equal means or medians and originate from the same population cannot be rejected. Considering the impact of the different adaptation conditions on the distributions of the intra-observer variability, significance can be inferred

Table 4.15 – Resulting p values and test statistics of the comparison of the inter- and intra-observer variability distributions in terms of both PF/3 and STRESS measures between the two different adaptation conditions at 3200 K and 5600 K ambient illumination. If not indicated otherwise (by footnotes) the applied statistical method was the standard two-sample t-test. Grey shaded table entries indicate significant differences between both adaptation conditions at a 5 % significance level. Additionally tabulated are the corresponding effect sizes in terms of Cohen's d.

	Inter-observer variability						Intra-observer variability					
	PF/3 measure			STRESS measure			PF/3 measure			STRESS measure		
	statistic	p value	Cohen's d	statistic	p value	Cohen's d	statistic	p value	Cohen's d	statistic	p value	Cohen's d
Asian Skin	0.584	0.564	0.217	0.333	0.741	0.122	0.634	0.531	0.232	1.417	0.168	0.517
Banana	−2.364*	0.027*	0.896*	−1.010	0.322	0.375	−1.427	0.165	0.541	−3.197*	0.005*	1.142*
Blueberry	0.138	0.891	0.052	0.089	0.930	0.033	0.108	0.915	0.039	−0.749	0.460	0.273
Blue Jeans	−2.562	0.016	0.936	−2.530	0.017	0.924	1.564	0.129	0.571	0.556	0.583	0.203
Broccoli	−1.668	0.107	0.609	−0.603	0.551	0.220	−2.727	0.011	1.013	−1.051	0.302	0.384
Butternut	−0.076	0.940	0.029	0.898	0.378	0.339	−0.207†	0.836†	0.078†	−0.072	0.944	0.027
Carrot	−0.570	0.573	0.208	−0.002	0.999	<0.001	0.290†	0.772†	0.106†	0.578	0.568	0.211
Caucasian	1.861	0.074	0.704	1.689	0.103	0.638	2.524	0.018	0.938	1.974	0.058	0.721
Concrete	−0.753	0.459	0.290	−1.342*	0.197*	0.526*	−2.403	0.024	0.926	−1.279	0.212	0.475
Green Salad	−1.488	0.150	0.585	−0.952†	0.341†	0.381†	−0.139	0.891	0.055	0.449	0.657	0.177
Red Cabbage	−0.386	0.702	0.141	−0.322	0.750	0.117	−0.049	0.961	0.018	0.317*	0.754*	0.116*
Red Rose	−1.282	0.211	0.477	−1.016	0.312	0.378	−0.450	0.656	0.167	−0.688	0.497	0.251

* Welch-test
† Wilcoxon rank-sum test

for the PF/3 data of the test objects of broccoli, Caucasian skin and concrete, whereas for the respective STRESS data only the test object of banana shows significant differences. For the rest of the test objects, the assumption of equal means or medians of the intra-observer variability data for the different adaptation conditions is met.

Regarding the calculated effect sizes, which are indicated in terms of Cohen's d (for further details on this measure see Sec. 4.4.2), small to moderate values are obtained for most test objects. A large effect size with $d > 0.8$, on the other hand, is found only for test objects where the overlap of the IQRs between the variability distributions of the PF/3 and STRESS measures obtained for the two different adaptation conditions shown in Figs. 4.27 and 4.28, respectively, is small compared to the respective absolute deviations in the median and/or mean variability values. In all of these cases, as can be seen from Table 4.15, statistical significance can be concluded.

However, with just a few test objects showing such significant differences between the observer variability distributions obtained for the two different adaptation conditions, the reported results must be concluded to be neither systematic nor an indicator for drawing universally valid conclusions about the general impact of the adapted white point on the observer variability in the present memory color rating experiments. Moreover, it seems that the observed variability in the subjects' ratings is more kind of an inherent, relatively stable feature of the corresponding experimental design and, therefore, less sensitive to changes in the adaptation conditions as one might expect at first glance, at least for the two different ambient illuminations at 3200 K and 5600 K considered in this thesis.

For cross-checking purposes, the overall distributions of the test objects' means and medians of the different variability measures should also be tested for significance. Starting with the inter-observer PF/3 measure the average mean (median) for all test objects with regard to the adaptation conditions at 3200 K and 5600 K is 31.69 (31.34) and 33.03 (32.70), respectively. For the intra-observer PF/3 values, one finds 17.64 (17.14) and 17.85 (17.64), whereas the inter-observer STRESS measure gives 0.2497 (0.2497) and 0.2551 (0.2538). Last but not least, the intra-observer STRESS values yield 0.1400 (0.1350) and 0.1402 (0.1409). In all of these cases, assessing the test objects at 3200 K ambient illumination slightly reduces the mean and

median inter- and intra-observer variability measures averaged over all familiar test objects. However, testing the underlying distributions for significance using again a two-sample t-test (necessary assumptions for being applicable here like normality and homoscedasticity were tested and confirmed first) revealed that the null hypothesis of equal sample means cannot be rejected adopting a 5 % significance interval and, therefore, no statistically significant differences can be concluded. This lack of significance in the impact of the adapted white point on the observer variability is consistent with the findings obtained from the comparisons of the individual variability distributions of each test object reported previously.

4.3.4 *Characteristics of Memory Colors Under Different Adaptation Conditions*

After having discussed the potential impact of the different adaptation conditions on the observer variability in the preceding section, a similar question arises regarding the influence of the observers' chromatic adaptation state on the fitted similarity distributions and the respective color centers of the various familiar test objects. With the color appearance rating experiments being conducted at two different ambient illuminations but apart from that at equal experimental conditions, the differences observed between both experiments and reported in Sec. 4.3.2 can most likely be attributed to the different spectral composition of the ambient illumination defining the white point of adaptation. However, the challenge now is to decide whether these differences are really statistically significant or occur simply by pure chance.

For this purpose, Figs 4.30 and 4.31 compare the contour line plots of the fitted similarity distribution functions of each test object in CIECAM02 chromaticity space as obtained for the two different adaptation conditions. As can be seen, slight deviations in the four ellipse dimensions shape, orientation, size, and location of the respective contour line plots are observed for all test objects. In some cases though they are more pronounced than in others without really showing any kind of systematic consitencies that could be attributed to the observers' state of chromatic adaptation. Consequently, again no general trend can be derived between the results of both experimental runs. In other words, no indication is given that a certain adaptation condition would alter the ellipse dimensions of the corresponding similarity distributions of all test objects in a general, well-defined way. Instead, a somewhat random pattern regarding the variations in shape, orientation, size, and location is observed. In this context, the most obvious nonconformities between the similarity distribution functions assigned to different CCTs are found for the test objects of Asian skin, banana, broccoli, Caucasian skin, green salad, red cabbage, and red rose, while for the remaining test objects the observed variations are less conspicuous.

Basically, all these findings mentioned above are in accordance to the previous work of Sanders [25] who with the results of his memory color rating experiments also reported non-systematic deviations of various degree between the tolerance ellipses fitted to his selection of test objects assessed under two different reference light sources. Hence, it can be concluded that the way the preferred color appearance of certain test objects is recalled by the observers while being adapted to a specific ambient illumination is more object-dependent rather than following a universal scheme. This means that for some of the test objects assessed in the current work, like e.g. blueberry, blue jeans, butternut squash,..., a quite good consistency in the color appearance ratings and, therefore, in the fitted similarity distributions is observed between the different adaptation conditions, while for the remaining test objects, where the manifestation of observed variations in the similarity distributions changes from object to object, no such memory color constancy could be inferred.

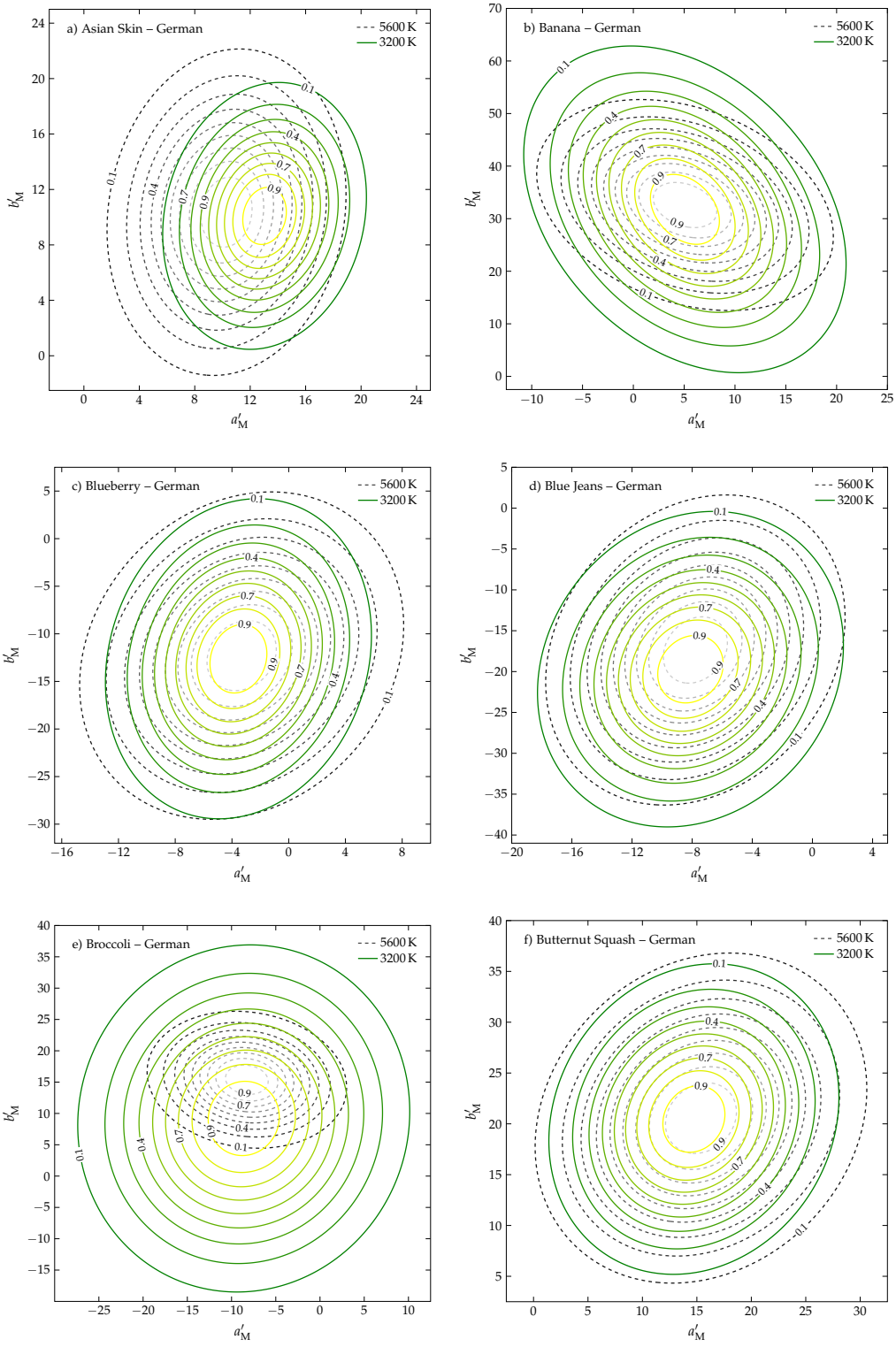


Figure 4.30 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the two different adaptation conditions at 3200 K (green-to-yellow colormap, solid line) and 5600 K (dark-to-light-gray colormap, dashed line) for the test objects of a) asian skin, b) banana, c) blueberry, d) blue jeans, e) broccoli, and f) butternut squash. By applying Box's *M*-test, significant differences in the covariance matrices defining the shape, size, and orientation of the similarity distributions can be found for the test objects of banana and broccoli. In addition, Hotelling's T^2-test reveals significant differences in the memory color centers between both adaptation conditions for the test objects of Asian skin and broccoli.

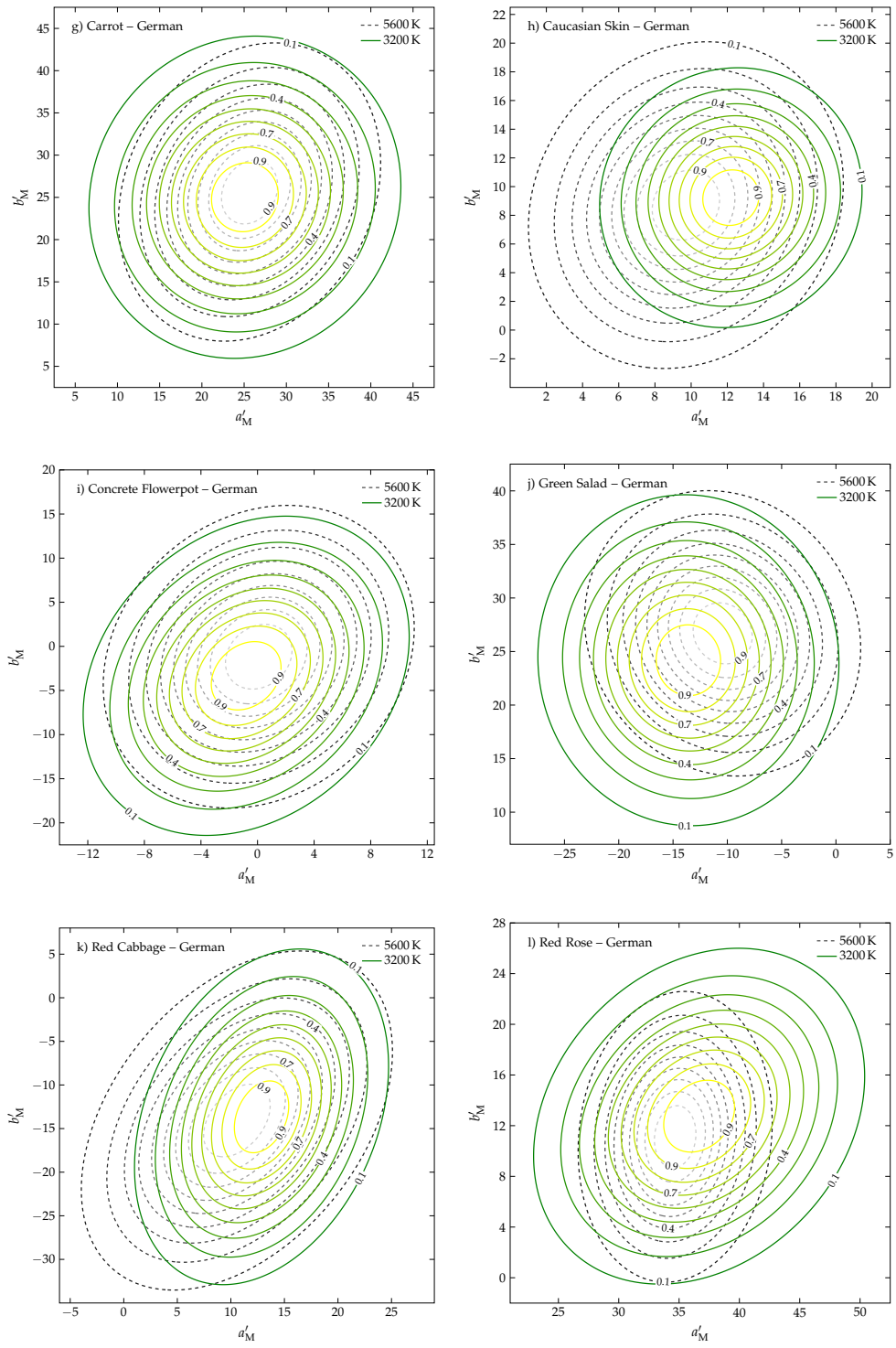


Figure 4.31 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the two different adaptation conditions at 3200 K (green-to-yellow colormap, solid line) and 5600 K (dark-to-light-gray colormap, dashed line) for the test objects of g) carrot, h) Caucasian skin, i) concrete flowerpot, j) green salad, k) red cabbage, and l) red rose. By applying Box's *M*-test, significant differences in the covariance matrices defining the shape, size, and orientation of the similarity distributions can be found for the test objects of Caucasian skin and red rose. In addition, Hotelling's T^2-test reveals significant differences in the memory color centers between both adaptation conditions for the test objects of Caucasian skin, green salad and red rose.

In order to determine whether or not these object-dependent deviations between the two adaptation conditions at 3200 K and 5600 K are significant, Box's M-test and Hotelling's T^2-test were again applied to compared the covariances and means of the underlying multivariate data samples. In Table 4.16 the corresponding results are summarized. As can be seen, the application of Box's M-test assuming a 5 % significance level revealed dealing with unequal covariances for the test objects of banana, broccoli, Caucasian skin, and red rose giving in each case a p value smaller than 0.04. From Figs. 4.30 and 4.31 it should be further noticed that for the test objects of banana and red rose these deviations in their corresponding covariance matrix are mainly due to differences in size and orientation of the respective similarity distribution functions, whereas for the test objects of broccoli and Caucasian skin also changes in shape can be observed.

Regarding the outcome of Hotelling's T^2-test, the null hypothesis of equal sample mean vectors and, therefore, of equal memory color centers must be rejected for the test objects of Asian skin, broccoli, Caucasian skin, green salad, and red rose. In all of these cases, the chromatic color differences reported in Sec. 4.3.2 between the memory color centers obtained for the two different adaptation conditions are considered to be statistically significant, which is further supported by the indicated, relatively large effect sizes D^2 (details on the calculation of the Mahalanobis distance D^2 are found in Sec. 4.4.3). For the remaining test objects no such significance in the mean vectors of their similarity distributions can be concluded. Nevertheless, it should be emphasized that with six out of twelve familiar test objects showing either significantly different covariance matrices defining the shape, size, and orientation of the corresponding similarity distribution functions or significantly different locations of the memory color centers in CIECAM02-UCS chromaticity space (or even both), distinctions should be made in the construction of an updated memory-based color quality metric between warm-white and rather cool-white test light sources in order to account for the noticeable impact of different adaptation conditions on the observers' memory color assessments.

Table 4.16 – Resulting p values and test statistics of Box's M-test and Hotelling's T^2-test applied to the similarity distribution functions of the twelve familiar test objects assessed under two different adaptation conditions at 3200 K and 5600 K ambient illumination to check for significant differences. If not indicated otherwise (by a footnote) the standard, homoscedastic version of Hotelling's T^2-test was used. Grey shaded table entries indicate significant differences between both adaptation conditions at a 5 % significance level. For each test object, the corresponding effect size is given in terms of the Mahalanobis distance D^2.

	Box's M-test		Hotelling's T^2-test		
	test statistic χ^2	p value	test statistic T^2	p value	effect size D^2
Asian Skin	6.645	0.0841	20.431	<0.0001	0.409
Banana	18.910	0.0003	0.720*	0.6977*	0.014*
Blueberry	3.811	0.2826	0.230	0.891	0.005
Blue Jeans	0.355	0.9493	3.526	0.1715	0.071
Broccoli	101.094	<0.0001	22.601*	<0.0001*	0.452*
Butternut	2.076	0.5568	0.487	0.7838	0.010
Carrot	5.124	0.1629	0.440	0.8024	0.009
Caucasian	8.487	0.0370	15.323*	0.0005*	0.306*
Concrete	0.593	0.8980	3.329	0.1893	0.067
Green Salad	3.507	0.3198	16.068	0.0003	0.321
Red Cabbage	4.143	0.2464	5.214	0.0737	0.104
Red Rose	28.244	<0.0001	8.931*	0.0115*	0.179*

* heteroscedastic version of Hotelling's T^2-test must be applied here

Table 4.17 – Results of the extra sum-of-squares F-test for the effect of the adapted white point/ambient illumination. Statistical significance is given for all twelve test objects. In addition, the corresponding effect size η^2 is tabulated indicating a large to medium statistical effect.

	test statistic F	$F_{\text{crit.}}(0.05, df_{\text{null}}, df_{\text{alt.}})$	df_{null}	$df_{\text{alt.}}$	p value	effect size η^2
Asian Skin	3.727	1.355	122	116	<0.0001	0.162
Banana	3.716	1.356	121	115	<0.0001	0.163
Blueberry	8.034	1.355	122	116	<0.0001	0.294
Blue Jeans	6.921	1.353	123	117	<0.0001	0.262
Broccoli	6.872	1.348	126	120	<0.0001	0.256
Butternut Squash	2.185	1.355	122	116	<0.0001	0.102
Carrot	3.248	1.355	122	116	<0.0001	0.144
Caucasian Skin	3.675	1.358	120	114	<0.0001	0.162
Concrete Flowerpot	6.841	1.351	124	118	<0.0001	0.258
Green Salad	1.857	1.358	120	114	0.0005	0.089
Red Cabbage	2.660	1.355	122	116	<0.0001	0.121
Red Rose	3.629	1.355	122	116	<0.0001	0.158

Another indication that one should not simply pool the data is given by the extra sum-of-squares F-test [55, 270, 271]. Basically, this test can be used to evaluate the goodness-of-fit of two alternative, nested models fitting the same data with one model being a simpler version (fewer parameters) of the other (more parameters). Based on statistical hypothesis testing, the extra sum-of-squares F-test compares the relative improvement in the sum-of-squares (SS) of the more complicated model with the relative loss of degrees of freedom that goes hand in hand. If the simpler model is correct (null hypothesis), the amount of improvement in the SS of the more complex model (alternative hypothesis) is assumed to be observed merely by chance, which in turn is determined by the degrees of freedom in each model. In other words, the extra sum-of-squares F-test compares the difference in the SS between booth models with the difference one would expect by chance, which can mathematically be expressed by

$$F = \frac{\left(\text{SS}_{\text{null}} - \text{SS}_{\text{alt.}}\right)/\text{SS}_{\text{alt.}}}{\left(df_{\text{null}} - df_{\text{alt.}}\right)/df_{\text{alt.}}} = \left(\frac{\text{SS}_{\text{null}}}{\text{SS}_{\text{alt.}}} - 1\right)\left(\frac{df_{\text{null}}}{df_{\text{alt.}}} - 1\right)^{-1}, \tag{4.21}$$

where SS_{null} and $\text{SS}_{\text{alt.}}$ are the residual sum-of-squares between the data to be fit and the respective model, while df_{null} and $df_{\text{alt.}}$ denote the corresponding degrees of freedom. Hence, the test statistic F equals the relative difference in the sum-of-squares between the simple and the more complex model divided by their relative difference in degrees of freedom.

In the present case of the color appearance rating experiments conducted at different CCTs of the ambient illumination, the simple model assumes that for each test object the variance of the entire rating data set obtained by pooling the respective results of both adaptation conditions can be explained by a single multivariate Gaussian function defined by seven fit parameters (see Sec. 4.2.4.2). The more complex model, on the other hand, postulates a separate Gaussian for each adaptation condition to explain the total variance in the rating data leading to a total number of 14 fit parameters, i.e., seven for each CCT. Based on these considerations, the null hypothesis assuming the correctness of the simple model can be evaluated for each familiar test object.

The corresponding results are summarized in Table. 4.17 and indicate a statistically significant effect for all twelve test objects ($p \leq 0.0005$) so that, adopting a 5 % significance level, the null hypothesis must be rejected in all case. In addition, the effect size $\eta^2 = \left(\text{SS}_{\text{null}} - \text{SS}_{\text{alt.}}\right)/\text{SS}_{\text{null}}$ ranges between 0.089 and 0.294 with an average value of 0.181 which, following Cohen's rule of thumb [272], is concluded to represent a medium to large statistical

effect. Hence, further evidence is given that pooling the observers' color appearance ratings obtained for different adaptation conditions would disregard a certain amount of important information and, therefore, lead to a false assumption about the characteristics of the memory color similarity distribution functions designated for constructing an updated memory-based color quality metric. Additional discussions regarding this issue are necessary and will be provided in the next chapter.

4.4 CROSS-CULTURAL VARIATIONS IN THE ASSESSMENT OF MEMORY COLORS

Apart from the impact of different adaptation conditions on the observers' memory color appearance ratings, several other potential influence factors on the perception of familiar objects could be identified (e.g., gender, age, expertise in color science,...). One of these factors that has in this specific context rarely been discussed in the literature is the cultural origin of the observers, which is supposed to have a significant impact on the memory color assessments in the current experimental design. For a further analysis of this anticipated effect, the experiments described in Secs. 4.2.2 and 4.3.1, which were conducted at different adaptation conditions, have been repeated for the same twelve familiar test objects and the same two ambient illumination settings with the originally German observers being replaced by Chinese subjects.

The reason here to focus on collecting the color appearance ratings of Chinese observers was basically twofold. First, the Chinese culture and traditions as well as the way how young people grow up in China are considered to be quite different from what is usually experienced here in Germany. Chinese people for example eat different food, they live in a different natural and architectural environment, they get a different education, etc.. For all these reasons and many more, it is assumed that Chinese people could have developed an own understanding of what is beauty and what is pleasant to them that might be clearly different from what a German observer would expect, which eventually would also lead to significant differences in the color appearance ratings of the present experiments. Second, with a huge Chinese community at the Technische Universität Darmstadt and some Chinese speaking staff at our institute, it was relatively easy and convenient to find enough volunteers to participate in the experiments. If having chosen another cultural target group, things might have gotten a little bit more difficult.

In total, 16 male and 18 female Chinese observers participated in the experiments. All of them were recruited among the university students showing again a varying degree of expertise in color science with ages ranging between 18 and 35 (∅ 25.1). Special care was taken to select the Chinese observers in such a way that their group composition regarding the aspects of age and experience in color science was approximately comparable to the group composition of the German observers to allow for a more reliable comparison between the results obtained for both cultural observer groups.

All Chinese observers were native Chinese which had not been born or grown up abroad. However, caution was required to further suppress induced cultural bias which might occur due to the process of adapting to the new cultural environment the Chinese observers were now living in since they had moved to Germany. For this reason and as far as possible, only those Chinese volunteers were selected to participate in the experiments who had been living in Germany for less than a year. Within this relative short period of time, no critical degree of cultural adaptation was expected. Unfortunately, with the additional constraints of age and experience in color science this selection criterion could not be complied for all test persons, i.e, two of the Chinese observers had been living in Germany for more than three years. Of

Table 4.18 – Mean values and corresponding ±1σ-intervals of the luminance measure and the CIECAM02-UCS lightness parameter for each test object as presented to the Chinese Observers for the 3200 K and 5600 K ambient illumination. Following the procedure introduced in Sec. 4.3.1.2 which allows for choosing test objects that exhibit similar reflectance characteristics and visual appearance as the objects used in the original experiments, the $RMSE_{adj.}$ and ΔE_{00} values are given to illustrate the deviations of the least variable object candidate finally selected to perform the Chinese experiments. In cases where non of these values is given, exactly the same test objects/hand models were used for all experiments. In addition, the max. $\Delta E'_{chrom.}$ values indicate the observed maximum CIECAM02-UCS chromatic differences between the objects' chromaticity coordinates shown in Figs. 4.16 and 4.23 that were presented to the German observers and the newly calculated object chromaticities being presented to the Chinese observers.

	3200 K adapted white point					5600 K adapted white point				
Test Object	$RMSE_{adj.}$	ΔE_{00}	Mean Luminance $\overline{L}$ in cd m^{-2}	Mean Lightness Parameter $\overline{J'}$	max. $\Delta E'_{chrom.}$	$RMSE_{adj.}$	ΔE_{00}	Mean Luminance $\overline{L}$ in cd m^{-2}	Mean Lightness Parameter $\overline{J'}$	max. $\Delta E'_{chrom.}$
Asian Skin	-	-	299.37 ± 5.98	78.54 ± 0.57	0.23	-	-	297.45 ± 5.87	76.32 ± 0.54	0.29
Banana	0.0168	1.34	500.19 ± 6.83	91.19 ± 0.58	0.50	0.0153	1.17	501.14 ± 7.02	91.39 ± 0.56	0.46
Blueberry	0.0128	0.93	94.13 ± 2.91	47.84 ± 0.52	0.37	0.0144	1.31	92.02 ± 2.24	47.11 ± 0.53	0.42
Blue Jeans	-	-	77.39 ± 2.58	43.50 ± 0.53	0.22	-	-	77.47 ± 2.30	43.76 ± 0.51	0.28
Broccoli	0.0107	1.96	103.63 ± 2.94	48.25 ± 0.50	0.45	0.0089	1.82	106.97 ± 2.98	49.37 ± 0.58	0.61
Butternut	0.0083	1.21	278.48 ± 6.36	74.06 ± 0.51	0.49	0.0118	1.59	265.40 ± 6.74	73.12 ± 0.45	0.38
Carrot	0.0115	0.87	288.53 ± 5.99	76.08 ± 0.48	0.36	0.0122	0.91	289.86 ± 6.35	76.09 ± 0.54	0.41
Caucasian	-	-	373.01 ± 6.10	83.30 ± 0.62	0.28	-	-	375.90 ± 6.39	83.57 ± 0.63	0.30
Concrete	-	-	368.85 ± 7.02	81.81 ± 0.58	0.67	-	-	377.46 ± 7.76	82.96 ± 0.55	0.56
Green Salad	0.0129	1.27	208.90 ± 5.31	64.82 ± 0.54	0.52	0.0157	1.62	202.01 ± 5.61	64.42 ± 0.56	0.33
Red Cabbage	0.0132	0.99	62.05 ± 1.89	39.77 ± 0.48	0.25	0.0101	0.73	64.71 ± 2.49	40.70 ± 0.58	0.27
Red Rose	0.0125	1.74	38.55 ± 0.91	33.29 ± 0.35	0.58	0.0089	1.24	36.78 ± 0.81	32.41 ± 0.46	0.45

course, it would be better to have the experiments entirely running with Chinese observers that have never been influenced by another culture before, but in the lack of the opportunity to bring the whole experimental setup to China and (re-)perform the experiments there, the current solution was the best that could be achieved for this thesis.

As in the experiments with the German observers, all Chinese participants were tested for normal or corrected-to-normal color vision. They received the same oral instructions as the German observers, which were simultaneously interpreted to Chinese language in case they did not understand the German explanation properly, and were again asked to rate the color appearance of the currently presented familiar test object according to their preference of how they thought the respective object should look like in reality adopting the same rating scale as used for gathering the German data. Like before, each of the twelve familiar test objects was rated for the two different adaptation settings by a total number of 15 individual Chinese observers that were for each combination of test object and adaptation condition more or less randomly selected from the bulk of all Chinese participants, where again an approximately equally balanced male-female ratio was targeted.

The Chinese experiments were conducted right after the German experiments were finished. Starting with the 3200 K adaptation condition, the experiments were again efficiently organized with three observers being tested at the same time so that all five test runs being necessary to obtain the complete rating data set for a specific test object could be completed within 2.5 hours. After finishing the experiments at 3200 K ambient illumination, the same was repeated for the 5600 K case. With two to three familiar objects being tested per week, the whole bunch of experiments for the Chinese observers could be finalized within a nine-weeks period.

During this whole period, regularly repeated measurements of the spectra of the ambient illuminations and the projector's basis spectra were again performed before and after each test session to check for the system's stability. From these measurements, similar excellent results as shown in Table 4.6 in terms of $\Delta u'v'$ color differences were obtained for the short-

and long-term stability of the illumination settings for the Chinese experiments. With no changes being made to the original experimental setup and illumination conditions, these measurement results also indicate that any kind of hardware aging effects, which might have occurred during the 21 weeks (but only ~260 hours of operation) the experiments took place in total, could be precluded so that no further fine-adjustment or recalibration of the lighting devices was necessary.

With the system's stability of the experimental setup being under control, the remaining challenge in preparing the color appearance ratings of the Chinese observers was again the proper selection of test object candidates whose spectral reflectance properties and visual characteristics resembled those of the test objects used in the original experiments as good as possible. In order to find such object candidates the same procedure as introduced in Sec. 4.3.1.2 was applied here. Table 4.18 summarizes for each test object the corresponding $\mathrm{RMSE}_{\mathrm{adj.}}$ and ΔE_{00} values of the least variable object candidate that was finally selected to perform the Chinese experiments. As can be seen, for all twelve familiar test objects these values were found to be sufficiently small indicating a reasonably good to excellent agreement between the reflectance spectra of the selected object candidate and its original counterpart.

With the reflectance and the projector's basis spectra being known, inverse mapping could then be used to calculate the RGB triplets with which the projector should be driven in order to precisely match for a given lightness the CIECAM02-UCS object chromaticities that had been presented to the German observers for the two different adaptation conditions (see Figs. 4.16 and 4.23, respectively). Hence, for each familiar test object, color appearance ratings of exactly the same chromatic stimuli were collected and could be compared for both cultural observer groups. In this context, the deviations between the original stimuli and the stimuli applied in the Chinese experiments are also expressed in Table 4.18 in terms of $\Delta E'_{\mathrm{chrom.}}$ values which, for each test object, indicate the observed maximum CIECAM02-UCS chromatic differences between the object's chromaticity coordinates as presented to the German observers and the newly calculated object chromaticities as presented to the Chinese observers for the two different ambient illumination conditions at 3200 K and 5600 K, respectively. As can be seen, for both illumination conditions these maximum chromatic

Table 4.19 – Average inter- and intra-observer PF/3 performance factors and STRESS values calculated from the color appearance ratings of the Chinese observers for the twelve familiar test objects at 3200 K and 5600 K adapted white point. In addition, the corresponding $\mathrm{ICC}(2,k)$ values are shown. Please note that for both measures PF/3 and STRESS the corresponding values of inter- and intra-observer variability are tabulated in the same column but with the latter given in parenthesis.

	3200 K adapted white point			5600 K adapted white point		
Test Object	Inter-(Intra-)observer PF/3	Inter-(Intra-)observer STRESS	$\mathrm{ICC}(2,k)$	Inter-(Intra-)observer PF/3	Inter-(Intra-)observer STRESS	$\mathrm{ICC}(2,k)$
Asian Skin	38 (22)	0.30 (0.18)	0.9016	38 (23)	0.29 (0.18)	0.9232
Banana	35 (17)	0.26 (0.14)	0.9429	35 (19)	0.26 (0.14)	0.9381
Blueberry	40 (21)	0.30 (0.16)	0.9010	43 (20)	0.33 (0.17)	0.8761
Blue Jeans	39 (20)	0.30 (0.16)	0.8848	37 (20)	0.29 (0.15)	0.9168
Broccoli	43 (17)	0.33 (0.14)	0.8606	35 (17)	0.27 (0.14)	0.9551
Butternut Squash	30 (17)	0.23 (0.13)	0.9587	37 (18)	0.28 (0.14)	0.8803
Carrot	34 (20)	0.25 (0.15)	0.9575	37 (21)	0.27 (0.15)	0.9324
Caucasian Skin	39 (21)	0.29 (0.17)	0.9070	33 (17)	0.25 (0.14)	0.9460
Concrete Flowerpot	39 (18)	0.30 (0.14)	0.9029	37 (18)	0.28 (0.14)	0.9151
Green Salad	33 (18)	0.25 (0.14)	0.9402	39 (21)	0.29 (0.15)	0.9150
Red Cabbage	35 (20)	0.27 (0.15)	0.9221	34 (21)	0.26 (0.16)	0.9433
Red Rose	32 (18)	0.24 (0.14)	0.9256	35 (19)	0.26 (0.15)	0.9327

differences range between 0.22 and 0.67 $\Delta E'_{\text{chrom.}}$ which in all cases is considered to be below the just noticeable difference threshold.

In addition, the obtained mean luminance and lightness values are also tabulated for each test object. When comparing these results with those reported for the German experiments in Tables 4.1 and 4.7, respectively, it can be concluded that excellent agreement within the $\pm 1\sigma$-intervals could be achieved for all test objects. Hence, any kind of potential impact of different lightness levels of the same test objects across the different experiments could most likely be precluded allowing for a reliable inter-cultural comparison between the chromaticity ratings of Chinese and German observers which will be reported in the following sections.

4.4.1 *Color Appearance Rating Results of Chinese Observers*

Before having a closer look on the impact of the cultural background on observer variability and memory color centers, the general color appearance rating results of the Chinese observers should be presented first. Starting again with the observed inter- and intra-observer variability to quantify the observers' rating precision and repeatability, Table 4.19 summarizes the corresponding PF/3, STRESS, and ICC$(2,k)$ values for the twelve familiar test objects assessed under both adaptation conditions. Regarding the PF/3 (STRESS) measure in case of the 3200 K adapted white point, values ranging from 30 (0.23) to 43 (0.33) with an average of 36.4 (0.28) for the inter-observer variability and from 17 (0.13) to 22 (0.18) with an average of 19.1 (0.15) for the intra-observer variability can be observed. For the 5600 K case, on the other hand, inter- and intra-observer PF/3 (STRESS) values ranging from 33 (0.25) to 43 (0.33) with an average of 36.7 (0.28) and from 17 (0.14) to 23 (0.18) with an average of 0.15 are obtained.

In comparison with the German variability results at 3200 K and 5600 K adapted white point, similar intra-observer PF/3 and STRESS values are noticed for the Chinese observer group. However, regarding the inter-observer variability measures, somewhat greater deviations are observed for most of the test objects with the Chinese observers showing larger PF/3 and STRESS measures and, therefore, slightly less precision in their memory color appearance ratings than their German counterparts. Especially for the test objects of broccoli and blueberry, a remarkably increase in the inter-observer PF/3 and STRESS values can be stated which indicates less familiarity with the respective objects leading to a larger variance in the color appearance ratings among different observers. This basically is in accordance to the fact that during the experiments a couple of Chinese observers reported difficulties in recalling from memory the typically associated object color for some of the test objects including broccoli and blueberry. With the object selection being based on an online survey conducted among German participants (even though other nationalities were not inherently excluded), it was expected that Chinese observers might have their difficulties with some of the test objects arising from a lack of familiarity. For this reason care must be taken when interpreting the results and attempting to derive general inter-cultural tendencies from the observers ratings.

The noticed larger variance in the color appearance ratings of the Chinese observers is also reflected by the tabulated ICC$(2,k)$ values which, for the majority of the test objects assessed under both adaptation conditions, are slightly smaller than those reported for the German observers. Basically, a minor ICC$(2,k)$ coefficient is the result of a larger variance in the underlying data which, therefore, cannot be well represented by simply assuming an average observer. Nevertheless, it must be stated here that in the present case the calculated ICC$(2,k)$ values for the Chinese observer group are still in a very good to excellent range so that, like for the German observers, an average Chinese observer can be postulated in

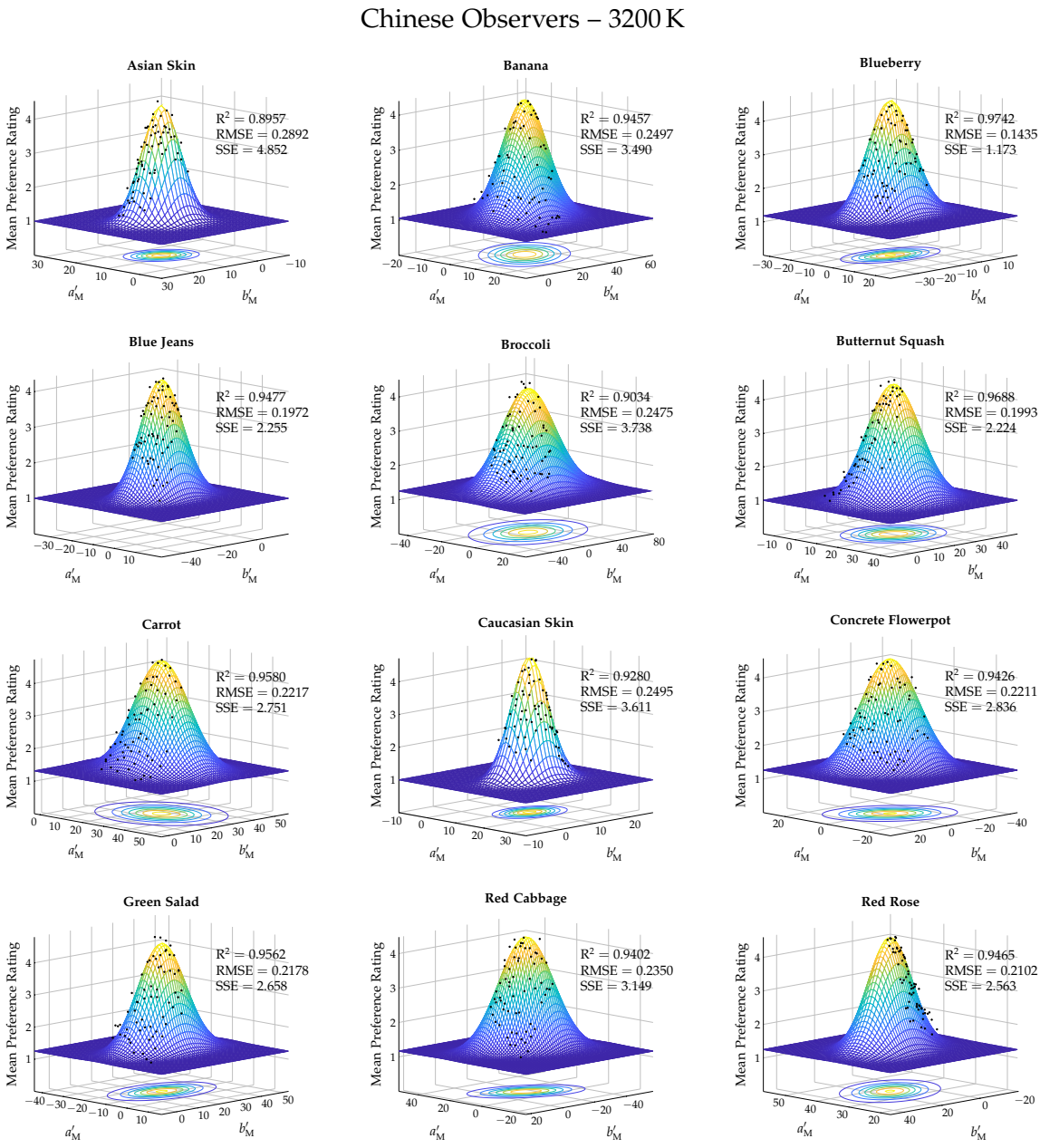


Figure 4.32 – Bivariate Gaussian similarity distributions fitted to the pooled Chinese observer data for each familiar test object modeling the mean preference ratings of an average Chinese observer in CIECAM02-UCS chromaticity space adapted to the 3200 K ambient illumination. Good to Excellent model performance based on the goodness-of-fit statistics can be concluded for all cases. Black dots represent the mean observer ratings for each projector setting, whereas the centroids of the Gaussian distributions define the chromaticity coordinates of the corresponding memory colors.

the following by pooling the individual observer ratings for each test object and adaptation condition.

Based on the pooled observer ratings, Gaussian modeling could eventually be applied to derive the memory color centers of an average Chinese observer for each of the twelve familiar test objects assessed under both adaptation conditions. The corresponding results are illustrated in Figs. 4.32 and 4.33, where, just as in the case of fitting the German data, the mean preference ratings of the average Chinese observer indicted by the black dots were modeled for each test object by a bivariate Gaussian distribution in CIECAM02-UCS (a'_M, b'_M) chromaticity space. From these model fits, the corresponding memory color centers were

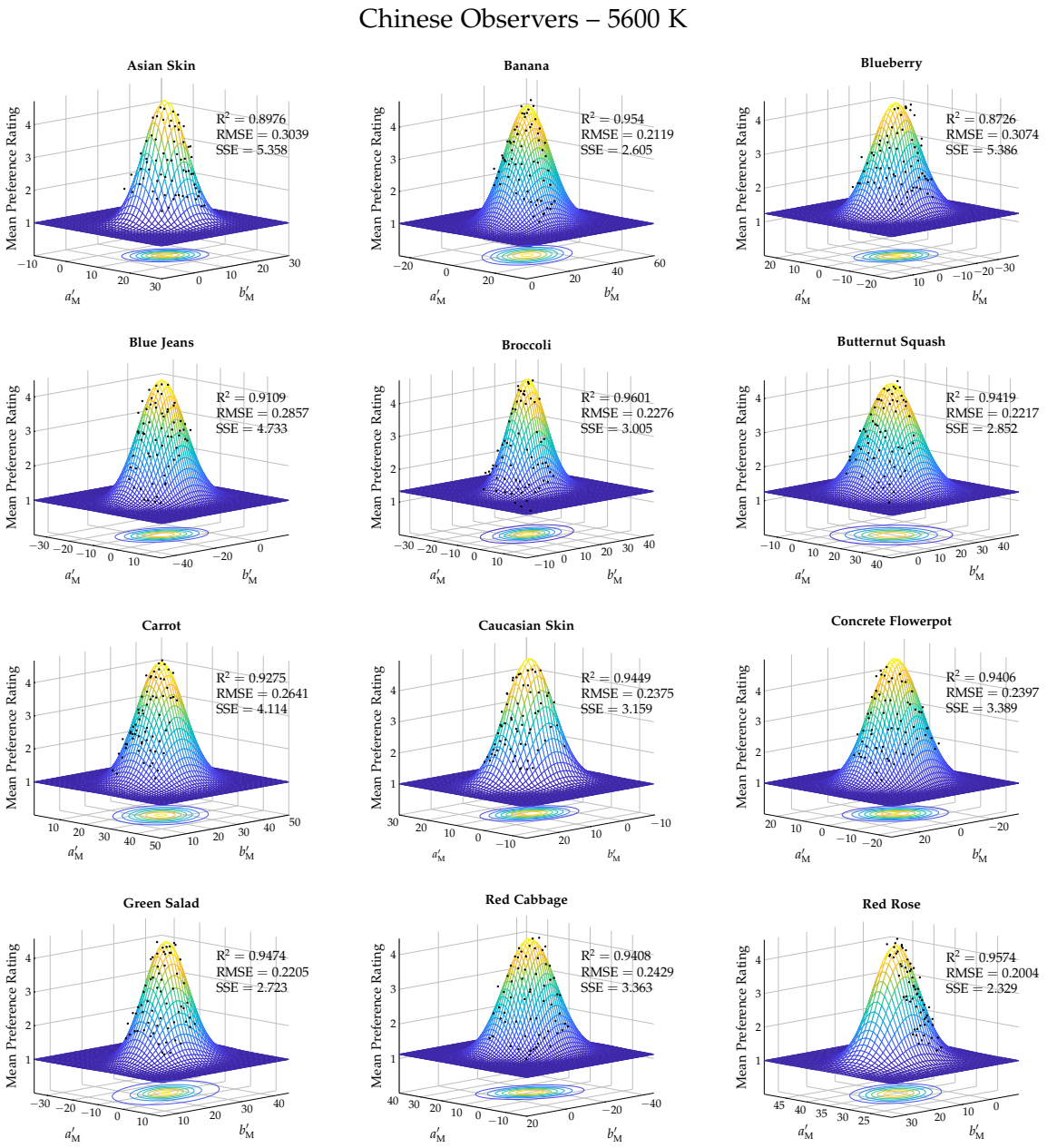


Figure 4.33 – Bivariate Gaussian similarity distributions fitted to the pooled Chinese observer data for each familiar test object modeling the mean preference ratings of an average Chinese observer in CIECAM02-UCS chromaticity space adapted to the 5600 K ambient illumination. Good to Excellent model performance based on the goodness-of-fit statistics can be concluded for all cases. Black dots represent the mean observer ratings for each projector setting, whereas the centroids of the Gaussian distributions define the chromaticity coordinates of the corresponding memory colors.

extracted as being given by the centroids of the Gaussian distributions and are summarized in Table 4.20 comparing the results of both adaptation conditions. In addition, the Pearson correlation coefficients ϱ between the average rating data and the model predictions are also tabulated. Mean values of 0.97 and 0.95 are obtained for the two different adapted white points at 3200 K and 5600 K, respectively. In combination with the reported coefficients of determination R^2, which in all cases indicate that more than 87.2 % of the total variations in the mean ratings of the pooled observer data is explained by the bivariate Gaussian distributions, excellent model performance can once again be concluded.

Table 4.20 – CIECAM02-UCS chromaticity coordinates a'_M and b'_M of the Chinese memory color centers which are obtained for the twelve familiar test objects assessed under both adaptation conditions at 3200 K and 5600 K ambient illumination. Additionally tabulated are the corresponding Pearson correlation coefficients ϱ describing the goodness-of-fit of the Gaussian modeling as well as the CIECAM02-UCS chromatic color differences $\Delta E'_{\text{chrom.}}$ between the Chinese and German results.

	3200 K adapted white point				5600 K adapted white point			
Test Object	a'_M	b'_M	Pearson ϱ	Chinese vs. German $\Delta E'_{\text{chrom.}}$	a'_M	b'_M	Pearson ϱ	Chinese vs. German $\Delta E'_{\text{chrom.}}$
Asian Skin	12.92	10.16	0.95	0.1368	10.29	10.86	0.95	0.5055
Banana	3.88	28.29	0.97	3.6811	4.44	30.83	0.94	1.8801
Blueberry	−2.85	−12.81	0.99	0.7257	−2.42	−10.54	0.93	1.9652
Blue Jeans	−7.63	−17.17	0.97	2.5863	−7.12	−14.75	0.95	2.6935
Broccoli	−9.62	14.38	0.95	5.2704	−8.30	16.10	0.98	0.7225
Butternut Squash	15.75	20.64	0.98	1.0431	14.19	19.70	0.97	1.4721
Carrot	27.49	26.63	0.98	2.8937	24.99	24.65	0.96	1.1821
Caucasian Skin	12.72	9.40	0.96	0.5580	8.46	8.25	0.97	1.3264
Concrete Flowerpot	−0.82	−3.54	0.97	0.2020	−0.48	−1.35	0.97	0.5925
Green Salad	−13.52	23.13	0.98	1.0566	−9.56	26.06	0.94	1.0864
Red Cabbage	11.91	−16.93	0.97	3.4152	10.41	−14.65	0.97	0.5893
Red Rose	37.11	13.21	0.97	0.6465	34.89	11.25	0.98	0.2278

By comparing the Chinese memory color centers with those obtained for the German observers summarized in Tables 4.3 and 4.9, relatively moderate deviations between these two cultural observer groups can be found. For the 3200 K ambient illumination, chromatic differences ranging between 0.14 and 5.27 $\Delta E'_{\text{chrom.}}$ with an average of only 1.85 $\Delta E'_{\text{chrom.}}$ are observed, whereas the 5600 K adapted white point yields chromatic differences ranging from 0.23 to 2.69 $\Delta E'_{\text{chrom.}}$ with an even smaller average of 1.19 $\Delta E'_{\text{chrom.}}$. It should be noted that in both cases the observed chromatic differences between the two cultural observer groups are slightly smaller than the average chromatic shift of the memory color centers induced by a change in the adaptation conditions which according to Sec. 4.3.2 is found to be of the order of ~2.3 $\Delta E'_{\text{chrom.}}$. Hence, it remains questionable if a significant impact of the cultural background on the color appearance ratings might be deduced based on the present data.

In order to better visualize the chromatic differences and tolerances in the average observer's ratings of both cultural observer groups, Figs. 4.34 and 4.35 individually compare the acceptance boundary ellipses for each of the twelve familiar test objects that were assessed by the two different observer groups at 3200 K and 5600 K ambient illumination, respectively. These acceptance boundaries were essentially calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3", which as stated earlier in this work is considered to be the just acceptable limit and, consequently, determines for each test object the chromatic tolerances observed in the current experiments. For convenience, the corresponding memory color centers are depicted as black crosses, while the dashed light-grey line segments illustrate the orientation of the ellipses' major semi-axes. The corresponding geometric ellipse parameters including the length of the major and minor semi-axes a and b as well as the measures of shape a/b, orientation θ, and size $A = \pi ab$ are summarized in Table 4.21.

As can be seen, the characteristics of the fitted acceptance boundaries are strongly object-dependent and to some extent governed by the white point of adaptation. In most cases, the corresponding contour line plots exhibit similar shape and orientation but distinct differences in size when comparing the results of both cultural observer groups. With the exceptions of Asian skin, carrot, and green salad, the German observer ratings tend to give smaller

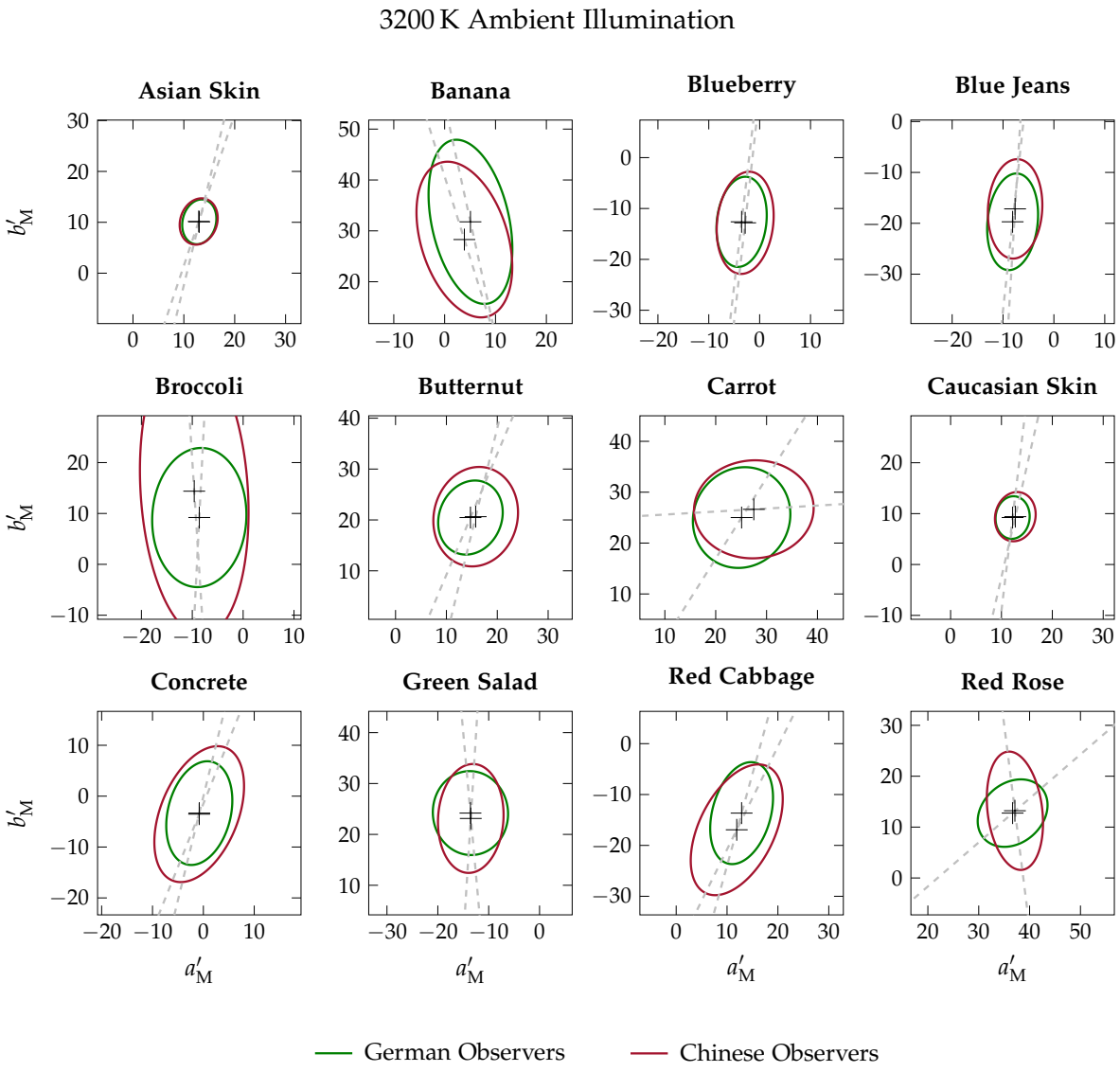


Figure 4.34 – Comparison of the acceptance boundary ellipses of the twelve familiar test objects between Chinese and German observers at 3200 K ambient illumination. These tolerance ellipses were calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3".

tolerance ellipses for both adaptation conditions. This indicates that on average Chinese observers are slightly more tolerant of and, at the same time, less sensitive to deviations from the objects' memory color centers compared to their German counterparts, which might also be due to the difficulties in recalling the typical object colors some of the Chinese participants reported to be confronted with during the experiments. Especially for the test object of broccoli assessed under the Planckian-like illumination a disproportionally large acceptance boundary is observed indicating a relatively huge unfamiliarity of the Chinese observers with the typical color appearance of the object when adapted to the 3200 K white point.

On a closer inspection and comparison of the tolerance ellipses of the different test objects, it can further be stated that in most cases the test objects exhibiting the most narrow Gaussian similarity distributions are those of Asian and Caucasian skin. This indicates that independent of their cultural background and the adapted white point observers are more sensitive to changes in the appearance of skin colors than to changes in the perceived chromaticities of

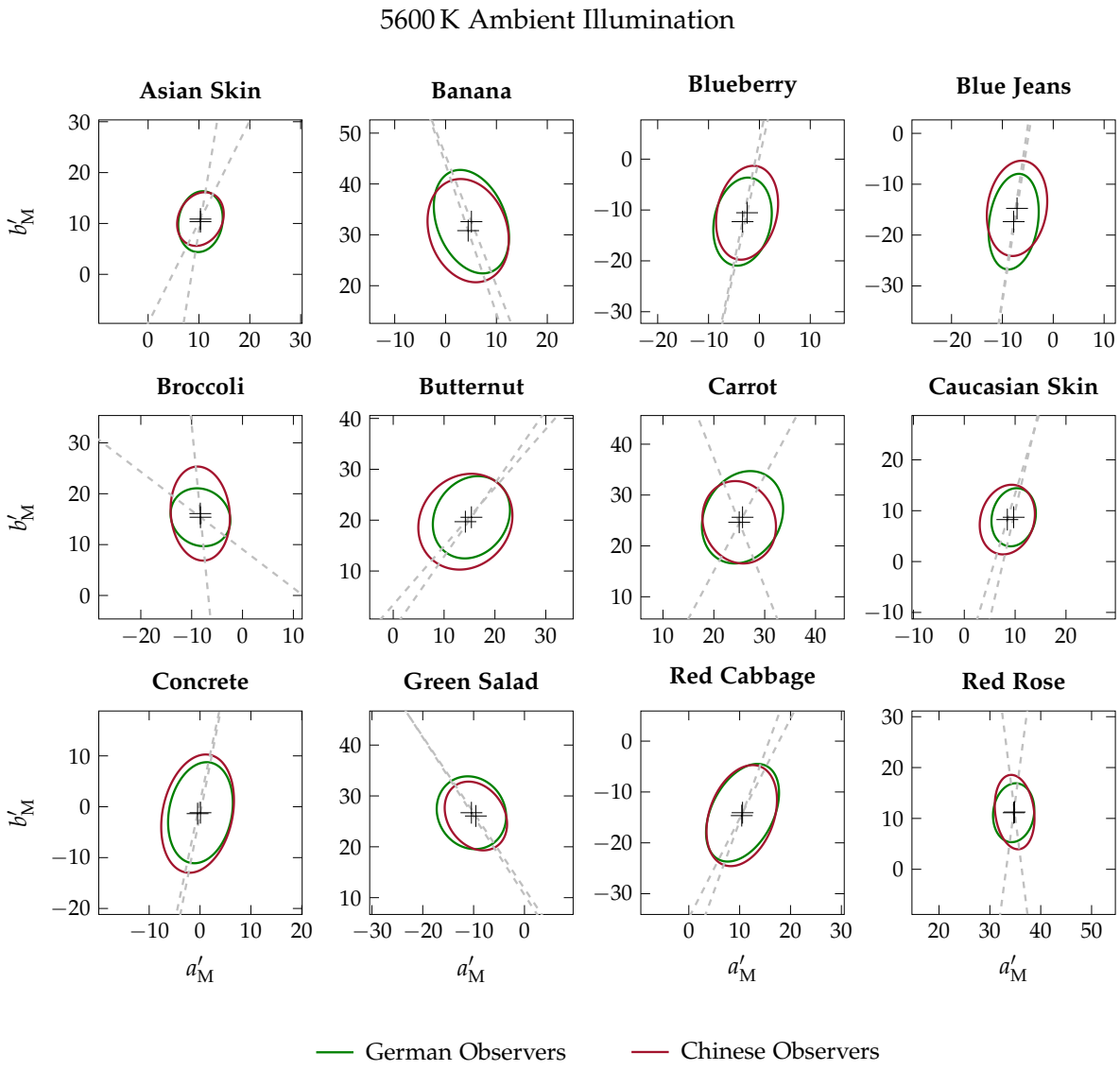


Figure 4.35 – Comparison of the acceptance boundary ellipses of the twelve familiar test objects between Chinese and German observers at 5600 K ambient illumination. These tolerance ellipses were calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3".

any of the remaining test objects. In addition, the corresponding tolerance ellipses obtained for both cultural observer groups show pronounced similarities not just in size but also in shape, orientation, and location. As a consequence, it can be concluded that people across various cultures and nationalities seem to have a quite common and consistent notion of how skin colors should ideally look like. This once again emphasizes the importance of skin colors in the context of memory and with regard to the intention of this thesis to eventually construct an updated, universally valid memory-based color quality metric.

Form Figs. 4.34 and 4.35, it should be further noted that despite the observed color differences of the memory color centers reported in Table 4.20, most of the familiar test objects generally show a large overlap between the chromatic tolerance ellipses of the two different observer groups. For the test objects of Asian skin, blueberry, broccoli, butternut squash, Caucasian skin, and concrete in case of the 3200 K ambient illumination and for the test objects of broccoli, butternut squash, Caucasian skin, and concrete regarding the 5600 K case, the

Table 4.21 – Overview of the chromatic differences and of the deviations in the perception correlates of chroma and hue between the memory color centers of the Chinese observers and the object chromaticities as perceived under the respective reference illumination. Additionally tabulated are the geometric measures of the corresponding observer tolerance ellipses. Please note that the semi-major axis a and semi-minor axis b are also given in units of $\Delta E'_{\text{chrom.}}$.

	3200 K adapted white point								5600 K adapted white point							
Test Object	$\Delta E'_{\text{chrom.}}$	ΔC	Δh in °	a	b	a/b	θ in °	πab	$\Delta E'_{\text{chrom.}}$	ΔC	Δh in °	a	b	a/b	θ in °	πab
Asian Skin	2.23	−2.97	−3.91	4.65	3.62	1.28	71.44	52.84	1.70	−2.46	−1.87	5.49	4.32	1.27	63.33	74.54
Banana	1.93	0.56	3.85	15.81	8.53	1.85	107.37	423.62	3.70	5.26	5.21	10.40	7.66	1.36	108.82	250.31
Blueberry	6.31	8.35	4.61	10.09	5.52	1.83	83.58	175.01	3.91	5.04	2.87	9.36	5.89	1.59	77.41	173.26
Blue Jeans	1.52	1.98	2.77	9.78	5.33	1.83	86.57	163.77	0.90	−1.40	0.01	9.44	5.82	1.62	81.06	172.64
Broccoli	4.83	6.35	8.81	28.46	10.54	2.70	93.63	942.80	5.96	8.02	10.12	9.27	5.79	1.60	95.54	168.72
Butternut Squash	4.54	8.25	−0.60	9.86	8.22	1.20	76.30	254.59	3.28	5.54	2.57	9.91	8.73	1.14	49.01	271.58
Carrot	6.83	15.80	2.00	11.77	9.63	1.22	3.37	355.89	4.83	9.27	4.38	8.25	7.06	1.17	111.51	182.89
Caucasian Skin	1.20	−0.64	−4.02	4.88	3.95	1.24	77.08	60.59	2.74	−3.67	−4.22	6.97	5.26	1.33	73.14	115.23
Concrete Flowerpot	6.69	0.64	182.39	14.07	7.63	1.84	68.07	337.30	4.55	−1.88	175.31	11.79	6.87	1.72	78.05	254.52
Green Salad	3.19	3.68	5.64	10.71	6.43	1.67	86.44	216.32	3.39	6.03	2.86	7.22	5.58	1.30	123.78	126.53
Red Cabbage	11.91	7.55	−35.01	13.86	7.41	1.87	63.69	322.47	7.38	6.21	−22.48	10.32	6.34	1.63	69.88	205.56
Red Rose	1.91	4.65	0.89	11.70	5.37	2.18	97.12	197.43	1.94	4.38	1.06	7.35	3.82	1.92	97.16	88.33

tolerance ellipses of the German observer group are even completely enclosed by the respective ellipses of the Chinese observer group. Based on these findings it is hardly surprising that the observed cross-cultural differences expressed in terms of $\Delta E'_{\text{chrom.}}$ values between the Chinese and German memory color centers are concluded to be much smaller than the extent of the overlap of the corresponding acceptance boundary ellipses.

Finally, the chromaticities of the Chinese memory color centers given in Table 4.20 should also be compared with the theoretical chromaticities the test objects would show when being solely illuminated by either of the two reference illuminants the ambient illumination of the present experiments was optimized for in order to provide adequate adaptation conditions. For this purpose, Figs. 4.36(a)-(b) illustrate the deviations between the objects' memory color centers of the Chinese observers and the corresponding reference chromaticities as perceived under the two different reference illuminants 3200 K Planckian and D56 using colored arrows which indicate both the direction and the amount of the observed memory-induced chromaticity shifts. In order to enable a better comparison, the results previously reported for the German observers are also (re-)plotted in Figs. 4.36(c)-(d).

Even though some slight differences in terms of ellipse location, size, shape, and orientation have been reported here, it can be seen that, with the exception of broccoli, the general pattern of the chromatic tolerance ellipses with respect to the observed memory-induced chromaticity shifts remains more or less the same when comparing the results of all four combinations of observer group and adaptation condition. As stated for the German observers, the chromaticity shifts of the Chinese subjects mainly point in the direction of increased CIECAM02-UCS colorfulness M' which due to proportionality also results in an increase in perceived chroma for most of the familiar test objects. From these increments in chroma ΔC, which are also summarized in Table 4.21 together with the corresponding hue shifts Δh and chromatic differences $\Delta E'_{\text{chrom.}}$ between the objects' reference and the memory color representations, it is concluded that the same general trend of memory colors to be shifted towards higher chroma values can also be derived from the Chinese rating data. On average, ΔC values of 4.52 and 3.36 are obtained for the two different adaptation conditions at 3200 K and 5600 K, which is approximately of the same order of magnitude as the average chroma increments obtained for the German observers given by 3.90 and 4.77, respectively.

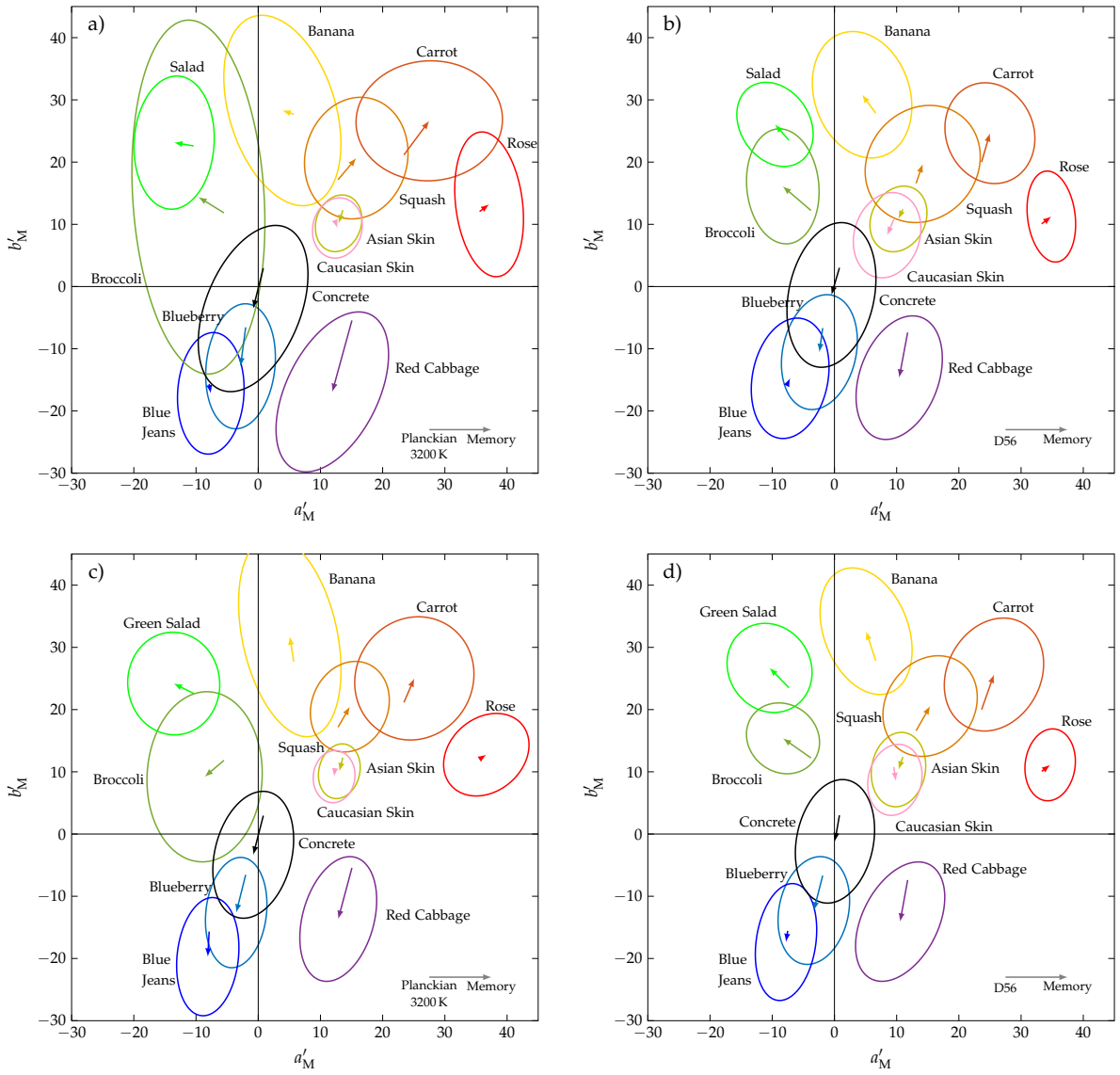


Figure 4.36 – Comparison between the memory color centers of the Chinese (upper row) and German (lower row) observers obtained for the two different adaptation conditions and the test objects' chromaticities rendered using the respective Planckian (left column) and daylight (right column) reference illuminants. Observed memory-induced chromaticity shifts are indicated by colored arrwos. Additionally shown are the acceptance boundary ellipses of each test object.

Similar findings can be reported for the hue shifts Δh and chromatic color differences $\Delta E'_{\text{chrom.}}$ calculated between the reference representations of the familiar test objects and their respective memory color centers. For most of the test objects assessed by the Chinese participants, the corresponding memory colors are again shifted towards the typical hues commonly associated with the respective objects. Besides being oriented in the same direction, these hue shifts also exhibit approximately the same magnitude as reported for the German observers, which can be seen by comparing Tables 4.4, 4.10, and 4.21. Regarding the chromatic color differences $\Delta E'_{\text{chrom.}}$ of the Chinese observers providing a measure for the strength of the long-term memory effects, average values of 4.42 $\Delta E'_{\text{chrom.}}$ and 3.69 $\Delta E'_{\text{chrom.}}$ can be extracted from Table 4.21 for the ambient illuminations of 3200 K and 5600 K, respectively. These average chromatic differences are in pretty good agreement to those found for the German observer group given by 4.16 $\Delta E'_{\text{chrom.}}$ and 4.19 $\Delta E'_{\text{chrom.}}$ so that the overall

strength of the long-term memory effects generally assumed to be observed in color appearance rating experiments of familiar objects can be estimated to be of the order of ~4 $\Delta E'_{\text{chrom.}}$.

Hence, with similar tendencies of the memory-induced chromaticity shifts being reported independently for each combination of cultural observer group and adaptation condition, it is likely that the observed trends derived from the present color appearance rating experiments and discussed above can be considered as an inherent feature of human color perception rather than being a matter of cultural peculiarity or adapted white point. Especially the general findings that memory colors of familiar objects in comparison to their appearance under reference illumination tend to be shifted towards higher chroma as well as towards their typical hues could be confirmed to be to some extent independent of the cultural background of the observers and of the light situation they are adapted to. In spite of observing the same tendencies regarding the memory-induced chromaticity shifts, a non-negligible statistical effect of the different adaptation conditions could however be reported in Sec. 4.3.4 for the color appearance rating results of the German observers. Consequently, it was first verified that such an effect also exists for the Chinese observers by repeating the same steps of analysis. Based on this verification, it had subsequently to be checked if the inter-cultural variations reported here in terms of deviations in observer variability and between the respective tolerance ellipses also show some significance or not. The corresponding considerations will be the focus of the following two sections.

4.4.2 *Statistical Inference on Observer Variability (II)*

Although the inspection of the different sets of acceptance boundary ellipses given in the preceding section revealed some general trends of agreement (orientation and magnitude of chromaticity shifts) and disagreement (ellipse parameters) in the assessment of memory colors between Chinese and German observers, the question still remains whether the observers' cultural background causes systematic and significant deviations between both observer groups. In the attempt of providing an adequate answer, the cultural impact on the various measures of observer variability should be examined first.

In order to be able to identify such inter-cultural effects, the observer variability distributions of both cultural observer groups in terms of PF/3 and STRESS units should be compared and tested for significance. Since an excellent correlation between the individual PF/3 and STRESS measures regarding both the inter- and intra-observer variability can also be stated for the Chinese subjects (Pearson's ϱ larger than 0.92 for all possible combinations, see Fig. 4.29), the following analysis will focus on the PF/3 measures only.

In Figs. 4.37 and 4.38, corresponding box plots visualize the distributions of the individual inter- and intra-observer PF/3 values of the Chinese observers as obtained for each test object assessed under both adaptation conditions at 3200 K and 5600 K, respectively. In addition, the differences in medians and means between Chinese and German observers were also calculated and are illustrated in the same figures on a second axis. As can be seen, the Chinese observers tend to show larger median and mean values for both inter- and intra-observer variability in comparison to the German results, which basically indicates slightly less precision and repeatability in the Chinese color appearance ratings. The largest deviations are found for the the inter-observer PF/3 distributions at 3200 K, where relatively speaking the Chinese subjects reported the most difficulties in recalling the typical object colors, whereas in the other cases less divergence is noticed. This general tendency of Chinese observers showing larger inter- and intra-observer variabilities, which gives a hint to less familiarity with the respective test objects, was also reported in Sec. 4.4.1. However, it must be noted

that for most of the test objects the overlap of the IQRs between the variability distributions of both cultural observer groups is found to be larger than the absolute difference in medians and/or means. This portends either a small to only moderate effect size or a lack of significance, which makes further analysis necessary.

At first, the ESD test is again applied to the various variability distributions of the Chinese observers in order to identify and remove clear outliers from the data sets. Subsequently, the corrected variability distributions are checked for normality using the KS and SW test as well as the z-scores of skewness and excess kurtosis as done in Sec. 4.3.3 for the German results. As can be seen from Tables 4.22 and 4.23, the assumption of dealing with normally

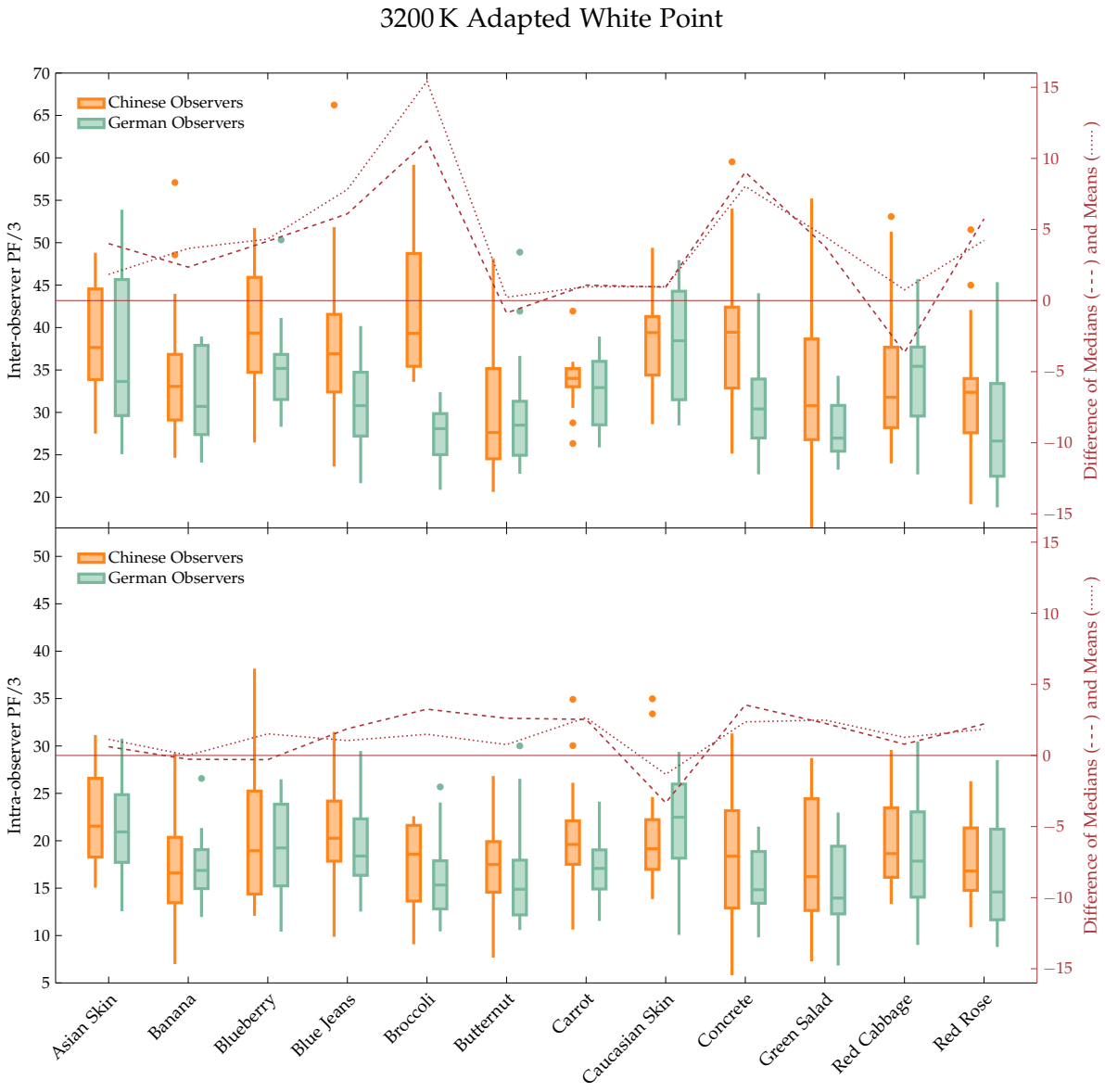


Figure 4.37 – Box plot comparison of the distributions of the individual inter-observer (upper plot) and intra-observer (lower plot) PF/3 values calculated for each combination of test object and adaptation condition at 3200 K. Here, the line inside the box gives the median of the distribution, while the bottom and top of the box represent its 25 % and 75 % quartile. The end of the whiskers indicate the lowest and highest data point lying at most 1.5 times the interquartile range below the lower and above the upper quartile, respectively. Any data point not included between the whiskers is identified as a suspected outlier and indicated by a colored dot. Additionally shown are the differences in medians (dashed line) and means (dotted line) between Chinese and German observers.

distributed Chinese inter- and intra-observer variability data holds, like in the case of the German observers (see Tables 4.11 and 4.12), for most of the familiar test objects assessed under both ambient illuminations. Inconsistencies are only observed for the test objects of broccoli and red cabbage regarding their inter-observer PF/3 distributions at 3200 K and for the test object of blueberry regarding its intra-observer variability at 5600 K. While in the case of red cabbage the corresponding variability distribution shows to much skewness for being normal, which is also confirmed by the outcome of the SW test, the shape of the inter-observer PF/3 distribution of broccoli is still acceptable. However, with both normality tests giving p values smaller than 0.011, the null hypothesis of normally distributed data

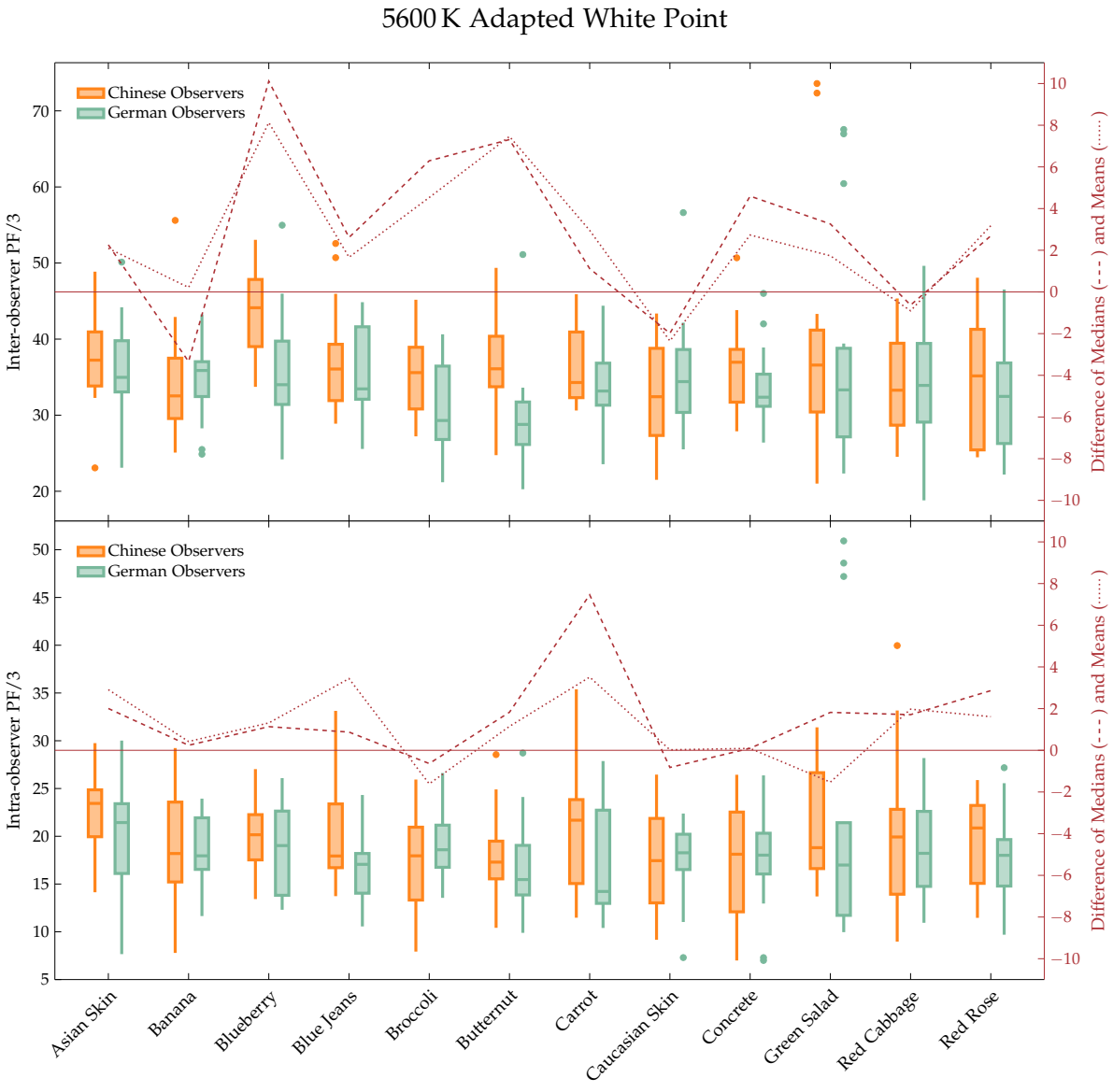


Figure 4.38 – Box plot comparison of the distributions of the individual inter-observer (upper plot) and intra-observer (lower plot) PF/3 values calculated for each combination of test object and adaptation condition at 5600 K. Here, the line inside the box gives the median of the distribution, while the bottom and top of the box represent its 25 % and 75 % quartile. The end of the whiskers indicate the lowest and highest data point lying at most 1.5 times the interquartile range below the lower and above the upper quartile, respectively. Any data point not included between the whiskers is identified as a suspected outlier and indicated by a colored dot. Additionally shown are the differences in medians (dashed line) and means (dotted line) between Chinese and German observers.

Table 4.22 – Results of various normality tests applied to the data distributions of the inter- and intra-observer PF/3 measure for German observers obtained at 3200 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Chinese inter-observer PF/3 at 3200 K						Chinese intra-observer PF/3 at 3200 K					
			KS[1] test		SW test				KS[1] test		SW test	
	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value
Asian Skin	0.359	−0.950	0.152	0.439	0.929	0.261	0.628	−0.997	0.139	>0.5	0.934	0.316
Banana	0.888	0.123	0.120	>0.5	0.955	0.581	0.457	−0.006	0.087	>0.5	0.986	0.996
Blueberry	0.155	−0.862	0.131	>0.5	0.966	0.803	1.661	0.177	0.185	0.178	0.908	0.112
Blue Jeans	0.627	0.449	0.136	>0.5	0.951	0.493	0.021	−0.223	0.109	>0.5	0.986	0.996
Broccoli	1.522	−0.651	0.252	0.011	0.823	0.007	−0.631	−1.084	0.194	0.130	0.900	0.095
Butternut	1.349	−0.139	0.180	0.208	0.924	0.219	−0.058	−0.585	0.118	>0.5	0.969	0.846
Carrot	−1.728	1.024	0.150	>0.5	0.890	0.104	−0.523	−0.413	0.220	0.083	0.893	0.106
Caucasian	0.174	−0.587	0.144	>0.5	0.970	0.870	0.310	−0.584	0.135	>0.5	0.960	0.778
Concrete	0.517	0.081	0.139	>0.5	0.953	0.513	0.167	−0.691	0.087	>0.5	0.987	0.997
Green Salad	0.960	0.297	0.191	0.144	0.949	0.426	0.432	−1.092	0.148	0.485	0.941	0.392
Red Cabbage	2.001	0.634	0.222	0.058	0.875	0.048	1.005	−0.648	0.181	0.201	0.926	0.235
Red Rose	−0.078	−0.175	0.185	0.257	0.947	0.547	0.546	−0.557	0.136	>0.5	0.974	0.908

[1] with Lilliefors correction [235]

Table 4.23 – Results of various normality tests applied to the data distributions of the inter- and intra-observer PF/3 measure for German observers obtained at 5600 K ambient illumination. In addition, the corresponding z-scores of skewness and excess kurtosis are listed. Grey shaded table entries shall indicate rejection of the null hypothesis of normal distributed data on a 5% significance level.

	Chinese inter-observer PF/3 at 5600 K						Chinese intra-observer PF/3 at 5600 K					
			KS[1] test		SW test				KS[1] test		SW test	
	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value	$z_{skewness}$	$z_{kurtosis}$	statistic	p value	statistic	p value
Asian Skin	1.203	−0.296	0.150	>0.5	0.911	0.165	−0.380	−0.334	0.172	0.262	0.943	0.428
Banana	0.611	−0.539	0.123	>0.5	0.975	0.933	0.009	−0.784	0.106	>0.5	0.961	0.717
Blueberry	0.164	−0.868	0.132	>0.5	0.960	0.698	0.451	−0.525	0.107	>0.5	0.964	0.766
Blueberry	1.122	0.211	0.152	>0.5	0.932	0.341	1.664	−0.075	0.236	0.033	0.881	0.061
Broccoli	0.293	−0.634	0.101	>0.5	0.975	0.921	−0.022	−0.712	0.096	>0.5	0.981	0.977
Butternut	0.051	−0.410	0.104	>0.5	0.985	0.993	0.287	−0.315	0.139	>0.5	0.958	0.687
Carrot	0.600	−1.132	0.214	0.060	0.912	0.144	0.972	−0.146	0.167	0.301	0.941	0.400
Caucasian	0.059	−0.922	0.123	>0.5	0.959	0.679	0.236	−0.955	0.151	0.454	0.960	0.691
Concrete	−0.255	−0.726	0.158	0.483	0.956	0.687	−0.322	−0.860	0.142	>0.5	0.955	0.603
Green Salad	−0.625	−0.333	0.174	0.337	0.948	0.570	0.969	−0.953	0.179	0.215	0.900	0.094
Red Cabbage	0.493	−0.968	0.124	>0.5	0.948	0.498	1.110	−0.168	0.166	0.352	0.929	0.295
Red Rose	0.029	−1.252	0.183	0.185	0.907	0.124	−0.353	−1.182	0.165	0.319	0.922	0.207

[1] with Lilliefors correction [235]

must clearly be rejected at a 5 % significance level. For the test object of blueberry, on the other hand, just the KS test shows a p value of less than this pre-specified significance level. Nevertheless, with a relatively large corresponding z-score of skewness being tabulated, even though it is slightly smaller than the limit of 1.96 (see Sec. 4.3.3), sufficient indication is given to reject the null hypothesis here.

In order to be able to draw statistically verified conclusions about potentially occurring systematic deviations between Chinese and German observers and, therefore, about the impact of inter-cultural variations on the observer variability when assessing memory colors, the proper statistical tests had to be chosen first. For this purpose, the BFL test was again applied to the Chinese and German variability data to check if homoscedasticity is fulfilled

Table 4.24 – Resulting p values and test statistics of the comparison of the inter- and intra-observer variability distributions in terms of both PF/3 and STRESS measures between the two different adaptation conditions at 3200 K and 5600 K ambient illumination. If not indicated otherwise (by footnotes) the applied statistical method was the standard two-sample t-test. Grey shaded table entries indicate significant differences between both adaptation conditions at a 5 % significance level. Additionally tabulated are the corresponding effect sizes in terms of Cohen's d.

	Inter-observer PF/3 variability						Intra-observer PF/3 variability					
	3200 K			5600 K			3200 K			5600 K		
	statistic	p value	Cohen's d	statistic	p value	Cohen's d	statistic	p value	Cohen's d	statistic	p value	Cohen's d
Asian Skin	0.601	0.553	0.219	2.002	0.056	0.314	0.622	0.539	0.227	1.571	0.127	0.574
Banana	0.486	0.631	0.508	−1.626	0.117	0.034	0.432	0.670	0.167	0.218	0.829	0.079
Blueberry	2.478*	0.022*	0.656	4.258	<0.001	1.151	0.656	0.517	0.240	0.747†	0.455†	0.275
Blue Jeans	2.677	0.012	0.989	−0.330	0.744	0.249	0.568	0.574	0.207	1.908	0.067	0.704
Broccoli	4.646†	<0.001†	3.202	2.107	0.044	0.769	1.434	0.163	0.534	−1.003	0.324	0.366
Butternut	1.071	0.294	0.031	4.306	<0.001	1.072	1.069†	0.285†	0.349	0.808	0.427	0.305
Carrot	1.108*	0.283*	0.249	1.496	0.146	0.546	0.466	0.645	0.175	1.410†	0.158†	0.520
Caucasian	0.409	0.686	0.154	−0.378	0.709	0.327	−2.077	0.048	0.804	−0.435	0.667	0.163
Concrete	2.592	0.015	1.045	2.031	0.053	0.491	1.119*	0.276*	0.412	−0.814	0.423	0.314
Green Salad	1.694*	0.108*	0.621	1.461	0.157	0.116	1.136	0.266	0.425	2.577	0.016	0.998
Red Cabbage	−0.677†	0.499†	0.101	−0.307	0.761	0.112	0.624	0.538	0.228	0.258	0.799	0.095
Red Rose	0.672	0.508	0.519	1.149	0.260	0.420	0.988	0.332	0.361	1.378	0.180	0.514

* Welch-test
† Wilcoxon rank-sum test

for each pair of distributions that should be compared. Here, the null hypothesis of equal variances must be rejected at a 5 % significance level for the inter-observer PF/3 variability distributions of blueberry ($F = 6.91$, $p = 0.014$), carrot ($F = 13.97$, $p < 0.001$), and green salad ($F = 4.58$, $p = 0.042$) at 3200 K ambient illumination and for the intra-observer PF/3 variability distribution of concrete ($F = 4.86$, $p = 0.036$) with the subjects being adapted to the same white point. For the remaining test objects at 3200 K as well as for all test objects assessed under 5600 K ambient illumination the null hypothesis can be confirmed and the assumption of equal variances between the respective variability distributions of the Chinese and German observers is consequently verified.

Hence, in cases where the assumptions of normality and homoscedasticity are met, the standard two-sample t-test was again applied to test for significant differences in the mean values of the variability distributions of both cultural observer groups. If the variability data are normal but the assumption of equal variances is violated, the Welch test like before was the method of choice. Furthermore, the Wilcoxon rank-sum test was once again used for comparing the Chinese and German variability distributions in cases where the data are not normally distributed for any of these two cultural observer groups. Table 4.24 finally summarizes for each test object assessed under both adaptation conditions the corresponding results of these statistical hypothesis tests.

As can been seen, significant differences which are indicated by grey shaded table entries between the variability distributions of Chinese and German observers are found for some of the test objects. In case of the inter-observer PF/3 measure, these are the test objects of blueberry, blue jeans, broccoli, and concrete at 3200 K as well as the test objects of blueberry, broccoli and butternut squash at 5600 K ambient illumination. Regarding the intra-observer PF/3 measure, significant differences are only observed for the test objects of Caucasian skin at 3200 K and green salad at 5600 K. For the remaining test objects, the null hypothesis of equal means or medians between the variability distribution of both cultural observer

groups cannot be rejected. In addition, the effect size is given in terms of Cohen's d [272–275] which reads

$$d = \frac{\mu_{\text{Chinese}} - \mu_{\text{German}}}{\sigma_{\text{pooled}}}, \tag{4.22}$$

where $\mu_{...}$ denotes the mean values of the variability distributions of the respective cultural observer group after outlier correction and σ_{pooled} is the pooled standard deviation given by

$$\sigma_{\text{pooled}} = \sqrt{\frac{(k_{\text{Chinese}} - 1)\,\sigma^2_{\text{Chinese}} + (k_{\text{German}} - 1)\,\sigma^2_{\text{German}}}{k_{\text{Chinese}} + k_{\text{German}} - 2}}, \tag{4.23}$$

where $k_{...}$ represents the sample size of the Chinese and German variability distributions that should be compared for a specific test object and $\sigma^2_{...}$ is the corresponding variance.

Please note that in cases where a non-parametric statistical test is applied for comparison, a correlation coefficient

$$r = \frac{Z}{\sqrt{k_{\text{Chinese}} + k_{\text{German}}}} \tag{4.24}$$

following the procedure suggested by Refs. [276, 277] was calculated first, where Z is the test statistic of the Wilcoxon rank-sum test. This correlation coefficient r could then be converted to Cohen's d using the equation

$$d = \frac{r\,(k_{\text{Chinese}} + k_{\text{German}})}{\sqrt{k_{\text{Chinese}} k_{\text{German}}\,(1 - r^2)}} \tag{4.25}$$

given by Rosenthal [278].

Basically, the effect size d describes the distance between the means (medians) of the observer variability distributions of Chinese and German observers in relation to the amount of the overlap between them. The larger Cohen's d, the easier detectable is the potential difference between two distributions. From Table 4.24 it can be concluded that in most cases a small ($d < 0.2$) to moderate effect size ($0.2 \leq d \leq 0.8$) according to Cohen's rule of thumb [272] is observed. As previously stated, there is substantial overlap in the corresponding distributions of the inter- and intra-observer PF/3 measures so that the null hypothesis of equal means and/or medians cannot be rejected.

A large effect size ($d > 0.8$), on the other hand, can be found for test objects where the overlap of the IQRs between the variability distributions of both cultural observer groups shown in Figs. 4.37 and 4.38 is small compared to the absolute difference in medians and/or means. In all of these cases, statistically significant differences between the compared variability samples must be reported. In combination with the large effect size this indicates a significant impact of the cultural background on the observer variability for these specific test objects. However, as stated previously this seems to be more a question of a lack of familiarity of the Chinese observers with the respective objects rather than being an inherent feature of intercultural variations in color appearance rating. With just a few test objects showing significant differences, the reported results are neither systematic nor allow for drawing universally valid conclusions.

Besides comparing the variability distributions of the Chinese and German observers separately for each test object, it is further of interest to have a look at the distributions of the overall mean and median inter- and intra-observer PF/3 measures. Regarding the Chinese observers, the average inter-observer PF/3 mean (median) value for all test objects is 35.58

(35.00) and 35.92 (35.66) at 3200 K and 5600 K ambient illumination, respectively. For the German observers average mean values of 31.69 (31.34) and 33.03 (32.70) can be found. Hence, it can be concluded that Chinese observers on average show clearly larger inter-observer PF/3 variability measures indicating less precision in their color appearance ratings. By running both two-sample t-test (assumptions of normality and homoscedasticity fulfilled for the 3200 K case) and Wilcoxon rank-sum test (assumption of normality violated for the distributions of mean and median PF/3 values of Chinese observers in the 5600 K case) on these data, the null hypothesis of equal sample means and/or medians must be rejected adopting a 5 % significance level. In all cases, a p value smaller than 0.03 and a relatively large effect size with $d > 0.94$ were obtained. Even though these results tempt to the conclusion of a significant impact of the observers' cultural background on the rating precision in the current experiments, one must be very careful with such an inference. Most likely, these findings are less a matter of inter-cultural variations than a question of test object familiarity. When for example excluding the possibly unfamiliar test objects, i.e., objects for which Chinese observers reported difficulties in recalling their typical object colors, from the statistical testing, no significant effect could be verified.

Regarding the average mean (median) of the intra-observer PF/3 measure, values of 18.77 (18.41) and 19.27 (19.23) are reported for the Chinese observers when being adapted to the 3200 K and 5600 K ambient illumination, respectively. For the German observers again somewhat smaller overall mean PF/3 values of 17.64 (17.14) and 17.85 (17.64) are observed. Running statistical testing, no significant differences could be found between the distributions of the mean and median intra-observer PF/3 values of both cultural observer groups for the 3200 K adapted white point ($p > 0.17$, $d < 0.6$). However, performing the experiment at 5600 K ambient illumination seemed to decrease the average repeatability in the observers' color appearance ratings for some of the test objects resulting in slightly larger PF/3 measures. This effect is more pronounced for the Chinese than for the German observers. Hence, by applying again a two-sample t-test to the underlying data distributions, significant differences between both cultural observer groups regarding the average mean and median intra-observer PF/3 measures must be stated ($p < 0.045$, $d > 0.87$). As can be seen from Fig. 4.38 this result is mainly based on an increasing relative difference between the mean (median) intra-observer PF/3 values of the Chinese and German observers for only some of the test objects like for example carrot or blue jeans. With no general systematic trend being observed, there is reason to believe that this finding is just an artifact of the relatively small sample size and, therefore, a result of pure chance.

4.4.3 *Impact of Cultural Background on Memory Color Appearance Ratings*

After having discussed the occurrence of cross-cultural variations in the inter- and intra-observer variability measures of the present color appearance rating experiments, a similar analysis should be performed in the following regarding the influence of the observers' cultural background on the fitted similarity distributions and the respective color centers of the various familiar test objects. With the color appearance rating experiments being conducted under more or less identical experimental conditions for both cultural observer groups, the differences observed between the Chinese and German subjects reported in Sec. 4.4.1 give indication that a cultural component in the assessment of memory colors may exist. Even though these differences are quite small and, therefore, as stated by Smet *et al.* [55], are likely to be of no practical importance, statistical analysis must be provided in order to get an idea about the significance and the size of the effect.

For this purpose, similar contour line representations as given in Figs. 4.30 and 4.31 could be drawn to compare the fitted similarity distribution functions of the Chinese and German observers for each test object. The corresponding plots are shown in Sec. A.2 of the Appendix. The general tendencies of the acceptance boundary ellipses, which can be considered as a two-dimensional projection of the fitted similarity distributions, of both cultural observer groups have already been discussed in Sec. 4.4.1 and should no be repeated here. Instead results of Box's M-test and Hotelling's T^2-test will directly be reported. In order to determine whether or not the observed inter-cultural deviations in the assessment of memory colors are significant, these statistical tests were the method of choice to compare the covariances and means of the bivariate similarity distributions that were fitted to the Chinese and German color appearance ratings gathered for both adaptation conditions.

Table 4.25 – Resulting p values and test statistics of Box's M-test and Hotelling's T^2-test applied to the similarity distribution functions of the twelve familiar test objects assessed under two different adaptation conditions at 3200 K and 5600 K ambient illumination to check for significant differences between Chinese and German observers. If not indicated otherwise (by a footnote) the standard, homoscedastic version of Hotelling's T^2-test was used. Grey shaded table entries indicate significant differences between both adaptation conditions at a 5 % significance level. Additionally tabulated are the corresponding effect sizes in terms of the Mahalanobis distance D^2.

	3200 K adapted white point					5600 K adapted white point				
	Box's M-test		Hotelling's t^2-test			Box's M-test		Hotelling's t^2-test		
	test statistic χ^2	p value	test statistic T^2	p value	effect size D^2	test statistic χ^2	p value	test statistic T^2	p value	effect size D^2
Asian Skin	0.354	0.950	0.053	0.974	0.001	2.732	0.435	0.699	0.705	0.014
Banana	2.939	0.401	5.539	0.063	0.111	1.679	0.642	2.191	0.335	0.044
Blueberry	3.483	0.323	1.229	0.541	0.025	0.139	0.987	2.602	0.272	0.052
Blue Jeans	2.358	0.502	3.314	0.191	0.066	6.504	0.090	3.974	0.137	0.079
Broccoli	56.945	<0.0001	1.537*	0.464*	0.031	17.652	0.001	0.306*	0.858*	0.006
Butternut	13.448	0.004	1.174*	0.556*	0.023	4.870	0.182	1.079	0.583	0.022
Carrot	3.120	0.374	4.075	0.130	0.081	7.393	0.060	1.432	0.489	0.029
Caucasian	0.449	0.930	0.964	0.618	0.019	2.465	0.482	2.749	0.253	0.055
Concrete	21.477	<0.0001	0.089*	0.957*	0.002	6.515	0.089	0.243	0.885	0.005
Green Salad	9.387	0.025	0.883*	0.643*	0.018	1.323	0.724	1.351	0.509	0.027
Red Cabbage	20.481	<0.0001	4.069*	0.131*	0.081	0.934	0.817	0.272	0.873	0.005
Red Rose	41.896	<0.0001	0.138*	0.933*	0.003	11.479	0.009	0.153*	0.926*	0.003

* heteroscedastic version of Hotelling's T^2-test must be applied here

In Table 4.25, the corresponding results are summarized. In addition to the various test statistics and p values, the multivariate effect sizes are also tabulated which are expressed in terms of the Mahalanobis distance D^2 [175] between the respective group means. In accordance to Cohen's d in the univariate case, the Mahalanobis distance gives a measure of the separation of the independent group means as a distance in space in relation to the pooled covariances of the two similarity distribution functions that should be compared. It can easily be shown (see e.g. Refs. [176, 279]) that

$$D^2 = \frac{(n_{\text{Chinese}} + n_{\text{German}})\, t^2}{n_{\text{Chinese}} n_{\text{German}}}, \tag{4.26}$$

where t^2 is the test statistic of Hotelling's T^2-test and n_{Chinese} and n_{German} are the sample sizes assumed for the bivariate similarity distributions of the Chinese and German observer group, respectively. In the present work, these sample sizes were both set to 100 and a multivariate random number generator [280] was used to calculate for each test object two independent data samples, one for the average Chinese and one for the average German observer, making use of the underlying fitted similarity distributions which are basically defined by their mean vectors and covariance matrices. These simulated data samples were finally used for running the statistical tests.

As expected from the discussions provided in Sec. 4.4.1, the null hypothesis of equal sample mean vectors could not be rejected for any of the familiar test objects assessed under both adaptation conditions. Thus, no significant differences in the memory color centers reported in Tables 4.3, 4.9, and 4.20 were found between Chinese and German observers. This is further emphasized by the calculated effect sizes which in all cases are quite small ($D^2 < 0.12$) indicating that for each test object the overlap of the fitted similarity distributions of both cultural observer groups was too large to be able to resolve the small difference between their mean vectors defining the respective memory color centers.

Nevertheless, significant differences between Chinese and German observers were obtained for the test objects of broccoli, butternut, concrete, green salad, red cabbage, and red rose at 3200 K and for the test objects of broccoli and red rose at 5600 K ambient illumination regarding the size, shape, and orientation of the corresponding similarity distribution functions given by their covariance matrices. Obviously, this is in accordance to the findings of Figs. 4.34 and 4.35, where due to the relatively large overlap between the Chinese and German acceptance boundary ellipses of these objects distinct deviations are only observed for the ellipse parameters of shape, size, and orientation but not for their locations. For all test objects, the observed cross-cultural differences expressed in terms of $\Delta E'_{\text{chrom.}}$ values between the respective memory color centers were generally found to be much smaller than the extent of the overlap of the corresponding acceptance boundary ellipses and/or similarity distributions.

Hence, the original claim of Smet *et al.* [55] that the impact of the cultural background of the observers on memory color assessments is negligibly small seems to be verified by the current results. However, it should be emphasized that even though no significant deviations can be found between the memory color centers of Chinese and German observers, it might be still important to take into account the differences in the covariance matrices of the fitted similarity distributions, which were reported for some of the test objects, when constructing an updated memory-based color quality metric that should be universally valid for Chinese as well as for German observers.

In order to get an idea whether this potentially inter-cultural effect shall be included or disregarded in the following, further analysis is necessary, which was again provided by applying the extra sum-of-squares F-test to the color appearance ratings of both cultural observer groups (see Sec. 4.3.4). For this purpose, an average global set of rating data was defined for each test object and adaptation condition by pooling the corresponding mean ratings of the Chinese and German observers. The simple model now assumes that the observed variance of this global set of rating data can be explained by a single bivariate Gaussian function defined by seven fit parameters (see Sec. 4.2.4.2). The more complex model, on the other hand, postulates a separate Gaussian for each observer group to explain the total variance in the combined set of rating data leading to a total number of 14 fit parameters. The null hypothesis assuming the correctness of the simple model was evaluated and the corresponding results are summarized in Table 4.26.

As can be seen, in most cases the null hypothesis must be rejected which indicates that the total variance in the rating data of the pooled average Chinese and German observers cannot be described by the simple model. Hence, it can be concluded that with the exceptions of butternut squash, carrot, and red rose at 3200 K and of Caucasian skin at 5600 K ambient illumination the extra sum-of-squares F-test revealed a non-negligible, statistically significant inter-cultural effect on the color appearance ratings between Chinese and German observers which is mainly caused by the differences in size, shape, and orientation of the respective sample distributions rather than by their centroid locations. However, the corresponding

Table 4.26 – Results of the extra sum-of-squares F-test for the effect of the observers' cultural background. Statistical significance could be confirmed for most but not for all of the twelve familiar test objects. On average, the corresponding effect sizes η^2 are about 40 % smaller than those obtained for the impact of the adapted white point.

	3200 K adapted white point			5600 K adapted white point		
	test statistic F	p value	effect size η^2	test statistic F	p value	effect size η^2
Asian Skin	4.709	<0.0001	0.196	4.825	<0.0001	0.200
Banana	3.199	<0.0001	0.144	2.004	<0.0001	0.094
Blueberry	2.363	<0.0001	0.109	2.159	<0.0001	0.098
Blue Jeans	2.989	<0.0001	0.132	1.761	0.0011	0.083
Broccoli	1.941	<0.0001	0.087	2.566	<0.0001	0.115
Butternut Squash	0.670	0.985	0.034	1.422	0.027	0.067
Carrot	0.374	0.998	0.019	2.418	<0.0001	0.112
Caucasian Skin	4.849	<0.0001	0.201	0.918	0.678	0.047
Concrete Flowerpot	1.870	0.0003	0.086	2.053	<0.0001	0.094
Green Salad	2.397	<0.0001	0.112	2.212	0.0005	0.104
Red Cabbage	1.923	0.0002	0.112	3.787	<0.0001	0.164
Red Rose	0.935	0.643	0.158	1.434	0.0256	0.069

effect sizes in terms of η^2 are quite small. On average, the effect size is approximately 40 % smaller than the effect size corresponding to the impact of the adapted white point (see Table 4.17) which gives an intrinsic hierarchy for the importance of these two different effects on the color appearance ratings of familiar objects considered in the current work. Even though inter-cultural variations are of minor importance compared to the influence of the chromatic adaptation conditions, both effects should initially be considered in the creation of an updated memory-based color quality metric for the color rendition evaluation of white light sources as described in the following chapter. A subsequent meta-correlation analysis will then be conducted to examine the predictive performance of such a metric in terms of visual appreciation of object colors rendered under various illuminations.

REFERENCES OF CHAPTER IV

[16] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.1, 1965.

[17] D. Nickerson and C. W. Jerome, "Color Rendering of Light Sources: CIE Method of Specification and its Application," *Illuminating Engineering*, vol. 60, no. 4, p. 262, 1965.

[19] W. Davis and Y. Ohno, "Color quality scale," *Optical Engineering*, vol. 49, no. 3, p. 033 602, 2010.

[20] M. S. Islam, R. Dangol, M. Hyvärinen, *et al.*, "Investigation of User Preferences for LED Lighting in Terms of Light Spectrum," *Lighting Research & Technology*, vol. 45, no. 6, pp. 641–665, 2013.

[23] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Memory Colours and Colour Quality Evaluation of Conventional and Solid-State Lamps," *Optics Express*, vol. 18, no. 25, pp. 26 229–26 244, 2010.

[24] G. B. Buck and H. C. Froelich, "Color Characteristics of Human Complexions," *Illuminating Engineering*, vol. 43, no. 1, p. 27, 1948.

[25] C. L. Sanders, "Color Preferences for Natural Objects," *Illuminating Engineering*, vol. 54, p. 452, 1959.

[26] C. L. Sanders, "Assessment of Color Rendition under an Illuminant Using Color Tolerances for Natural Objects," *Illuminating Engineering*, vol. 54, p. 640, 1959.

[27] D. B. Judd, "A Flattery Index for Artificial Illuminants," *Illuminating Engineering*, vol. 62, p. 593, 1967.

[30] W. A. Thornton, "A Validation of the Color-Preference Index," *Journal of the Illuminating Engineering Society*, vol. 4, no. 1, p. 48, 1974.

[31] P. Siple and R. M. Springer, "Memory and Preference for the Colors of Objects," *Perception & Psychophysics*, vol. 34, no. 4, p. 363, 1983.

[32] E. Hering, *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer Verlag, 1920.

[33] C. J. Bartleson, "Color in Memory in Relation to Photographic Reproduction," *Photographic Science and Engineering*, vol. 5, no. 6, p. 327, 1961.

[34] S. M. Newhall, R. W. Burnham, and J. R. Clark, "Comparison of Successive with Simultaneous Color Matching," *Journal of the Optical Society of America*, vol. 47, no. 1, p. 43, 1957.

[35] C. J. Bartleson, "Memory Colors of Familiar Objects," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 73, 1960.

[36] C. D. Hendley and S. Hecht, "The Colors of Natural Objects and Terrains, and their Relation to Visual Color Deficiency," *Journal of the Optical Society of America*, vol. 39, no. 10, p. 870, 1949.

[37] J. Pérez-Carpinell, M. D. de Fez, R. Baldoví, *et al.*, "Familiar Objects and Memory Color," *Color Research & Application*, vol. 23, no. 6, pp. 416–427, 1998.

[38] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Representation of Memory Prototype for an Object Color," *Color Research & Application*, vol. 24, no. 6, pp. 393–410, 1999.

[39] P. Bodrogi and T. Tarczali, "Colour Memory for Various Sky, Skin, and Plant Colours: Effect of the Image context," *Color Research & Application*, vol. 26, no. 4, p. 278, 2001.

[40] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Colour Appearance Rating of Familiar Real Objects," *Color Research & Application*, vol. 36, no. 3, pp. 192–200, 2011.

[41] T. Tarczali, S. Du Park, P. Bodrogi, *et al.*, "Long-term Memory Colors of Korean and Hungarian Observers," *Color Research & Application*, vol. 31, no. 3, p. 176, 2006.

[42] A. C. Hurlbert and Y. Ling, "If It's a Banana, It must be Yellow: The Role of Memory Colors in Color Constancy," *Journal of Vision*, vol. 5, no. 8, p. 787, 2005.

[43] T. Hansen, M. Olkkonen, S. Walter, *et al.*, "Memory Modulates Color Appearance," *Nature Neuroscience*, vol. 9, no. 11, p. 1367, 2006.

[44] J. J. M. Granzier and K. R. Gegenfurtner, "Effects of Memory Color on Color Constancy for Unknown Colored Objects," *i-Perception*, vol. 3, no. 3, p. 190, 2012.

[45] P. Bodrogi and T. Tarczali, "Investigation of Colour Memory," in *Colour Image Science: Exploiting Digital Media*, L. W. MacDonald and M. R. Luo, Eds. Chichester: John Wiley & Sons Limited, 2002, pp. 23–48.

[46] H. Zeng and R. Luo, "Modelling Memory Color Region for Preference Color Reproduction," in *Proceedings of the SPIE 7528, Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*, San Jose, CA, USA: International Society for Optics and Photonics, 2010, p. 752 808.

[47] S. Xue, M. Tan, A. McNamara, *et al.*, "Exploring the Use of Memory Colors for Image Enhancement," in *Proceedings of the SPIE 9014, Human Vision and Electronic Imaging XIX*, San Francisco, CA, USA: International Society for Optics and Photonics, 2014, p. 901 411.

[48] C. Boust, H. Chahine, M. B. Chouikha, *et al.*, "Color Correction Judgements of Digital Images by Experts and Naive Observer," in *Proceedings of the PICS Conference 2003*, Rochester, NY, USA: Society for Imaging Science and Technology (IS&T), 2003, pp. 4–9.

[49] C. Boust, H. Brettel, F. Viénot, *et al.*, "Color Enhancement of Digital Images by Experts and Preference Judgements by Observers," *Journal of Imaging Science and Technology*, vol. 50, no. 1, pp. 1–11, 2006.

[50] C. Boust, F. Cittadini, M. B. Chouikha, *et al.*, "Does an Expert Use Memory Colors to Adjust Images?" In *Proceedings of the 12th Color and Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 2004, pp. 347–353.

[51] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Color Reproduction and the Naturalness Constraint," *Color Research & Application*, vol. 24, no. 1, pp. 52–67, 1999.

[52] K. Smet, S. Jost-Boissard, W. R. Ryckaert, *et al.*, "Validation of a Colour Rendering Index Based on Memory Colours," in *Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2010, pp. 136–142.

[53] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "A Memory Colour Quality Metric for White Light Sources," *Energy and Buildings*, vol. 49, p. 216, 2012.

[54] K. Smet and P. Hanselaer, "Memory and Preferred Colours and the Colour Rendition of White Light Sources," *Lighting Research & Technology*, vol. 48, no. 4, p. 393, 2016.

[55] K. A. G. Smet, Y. Lin, B. V. Nagy, *et al.*, "Cross-Cultural Variation of Memory Colors of Familiar Objects," *Optics Express*, vol. 22, no. 26, p. 32 308, 2014.

[97] M. R. Luo and C. Li, "CIECAM02 and Its Recent Developments," in *Advanced Color Image Processing and Analysis*, C. Fernandez-Maloigne, Ed. New York, USA: Springer Science+Business Media, 2013, pp. 19–58.

[152] Commission Internationale de l'Éclairage, "Colour Fidelity Index for Accurate Scientific Use," *CIE Technical Report 224*, 2017.

[153] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Correlation Between Colour Quality Metric Predictions and Visual Appreciation of Light Sources," *Optics Express*, vol. 19, no. 9, pp. 8151–8166, 2011.

[158] C. I. de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.2, 1974.

[175] P. C. Mahalanobis, "On the Generalised Distance in Statistics," *Proceedings of the National Institute of Science in India*, vol. 2, no. 1, pp. 49–55, 1936.

[176] M. Sapp, F. Obiakor, A. J. Gregas, *et al.*, "Mahalanobis Distance: A Multivariate Measure of Effect in Hypnosis Research," *Sleep and Hypnosis*, vol. 9, no. 2, pp. 67–70, 2007.

[177] F. Ebner, "Derivation and Modelling Hue Uniformity and Development of the IPT Color Space," PhD thesis, Rochester Institute of Technology, 1998.

[178] F. Ebner and M. D. Fairchild, "Development and Testing of a Color Space (IPT) with Improved Hue Uniformity," in *Proceedings of the 6th Color and Imaging Conference: Color Science, Systems and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 1998, pp. 8–13.

[181] K. Duncker, "The Influence of Past Experience upon Perceptional Properties," *American Journal of Psychology*, vol. 52, no. 2, pp. 255–265, 1939.

[182] J. S. Bruner, L. Postman, and J. Rodrigues, "Expectation and the Perception of Color," *American Journal of Psychology*, vol. 64, no. 2, pp. 216–227, 1951.

[183] R. S. Harper, "The Perceptual Modification of Colored Figures," *American Journal of Psychology*, vol. 66, no. 1, pp. 86–89, 1953.

[184] C. Witzel, H. Valkova, T. Hansen, *et al.*, "Object Knowledge Modulates Colour Appearance," *i-Perception*, vol. 2, no. 1, pp. 13–49, 2011.

[185] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, "Color Appearance of Familiar Objects: Effetcs of Shape, Texture, and Illumination Changes," *Journal of Vision*, vol. 8, no. 5, pp. 1–16, 2008.

[186] Y. Ling and A. C. Hurlbert, "Color and Size Interactions in a Real 3D Object Similarity Task," *Journal of Vision*, vol. 4, no. 9, pp. 721–734, 2004.

[187] R. S. Berns and M. E. Gorzynski, "Simulating Surface Colors on CRT Displays: The Importance of Cognitive Clues," in *Proceedings of the AIC Midterm Meeting: Colour and Light*, Sydney, Australia: Association Internationale de la Couleur (AIC), 1991, pp. 21–24.

[188] D. H. Brainard, "Color Constancy in the Nearly Natural Image Part 2: Achromatic Loci," *Journal of the Optical Society of America A*, vol. 15, no. 2, pp. 307–325, 1998.

[189] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 5–55, 1932.

[190] P. Bodrogi and T. Q. Khanh, *Illumination, Color and Imaging: Evaluation and Optimization of Visual Displays*, T. Lowe, Ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2012.

[191] Y. Zhu, M. R. Luo, S. Fischer, *et al.*, "Long-term Memory Color Investigation: Culture Effect and Experimental Setting Factors," *Journal of the Optical Society of America A*, vol. 34, no. 10, pp. 1757–1768, 2017.

[192] S. Fischer, P. Bodrogi, T. Q. Khanh, *et al.*, "Memory Colors Part I: Comparison of Chinese and German Observers," in *Proceedings of the LICHT Symposium 2016*, Karlsruhe: Karlsruher Institut für Technologie (KIT), 2016, pp. 102–110.

[193] Y. Zhu, M. Luo, L. Xu, *et al.*, "Investigation of Memory Colours Across Cultures," in *Proceedings of the 23rd Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2015, pp. 133–136.

[194] W. Davis and Y. Ohno, "Toward an Improved Color Rendering Metric," in *Proceedings of the SPIE 5941, 5th International Conference on Solid State Lighting*, Bellingham, WA, USA: International Society for Optics and Photonics, 2005, 59411G–1–59411G–8.

[195] S. Jost-Boissard, M. Fontoynont, and J. Blanc-Gonnet, "Perceived Lighting Quality of LED Sources for the Presentation of Fruit and Vegetables," *Journal of Modern Optics*, vol. 56, no. 13, pp. 1420–1432, 2009.

[196] M. Royer, A. Wilkerson, M. Wei, *et al.*, "Human Perceptions of Colour Rendition Vary with Average Fidelity, Average Gamut, and Gamut Shape," *Lighting Research & Technology*, vol. 49, no. 8, pp. 966–991, 2017.

[197] A. Werner, "Spatial and Temporal Aspects of Chromatic Adaptation and their Functional Significance for Colour Constancy," *Vision Research*, vol. 104, pp. 80–89, 2014.

[198] G. Derefeldt, T. Swartling, U. Berggrund, *et al.*, "Cognitive Color," *Color Research and Application*, vol. 29, no. 1, pp. 7–19, 2004.

[199] O. Rinner and K. R. Gegenfurtner, "Time Course of Chromatic Adaptation for Color Appearance and Discrimination," *Vision Research*, vol. 40, no. 14, pp. 1813–1826, 2000.

[200] S. Ishihara, *Ishihara's Tests for Colour Deficiency: 38 Plates Edition*. Tokyo: Kanehara Trading Inc., 2016.

[201] H. Ichikawa, K. Hukami, and S. Tanabe, *Standard Pseudoisochromatic Plates Part II: For Acquired Color Vision Defects*. Tokyo, New York: Igaku-Shoin, 1983.

[202] A. Linksz, "The Farnsworth Panel D-15 Test," *American Journal of Ophthalmology*, vol. 62, no. 1, pp. 27–37, 1966.

[203] Richmond Products Inc., *Farnsworth D-15 and Lanthony Test Instructions*, `http://www.richmondproducts.com/files/8113/1550/0538/FR_15_Farnsworth_and_LanthonyD15_Instructions_Rev_1.7_0506.pdf`, [Online; accessed 07-August-2017], 2006.

[204] S.-S. Guan and M. R. Luo, "Investigation of Parametric Effects Using Small Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 331–343, 1999.

[205] J. B. Kruskal, "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[206] A. P. M. Coxon, *The User's Guide to Multidimensional Scaling: With Special Reference to the MDS(X) Library of Computer Programs*. Portsmouth, NH, USA: Heinemann Educational Publishers, 1982.

[207] E. Coates, K. Y. Fong, and B. Rigg, "Uniform Lightness Scales," *Journal of the Society of Dyers and Colourists*, vol. 97, no. 4, pp. 179–183, 1981.

[208] S. Adler, K. P. Chaing, T. F. Chong, *et al.*, "Uniform Chromaticity Scales: New Experimental Data," *Journal of the Society of Dyers and Colourists*, vol. 98, no. 1, pp. 14–20, 1982.

[209] W. Schultze, "The Usefulness of Colour-Difference Formulae for Fixing Colour Tolerances," in *Proceedings of the Helmholtz Memorial Symposium on Color Metrics*, Soesterberg, Holland: Association Internationale de la Couleur (AIC), 1972, pp. 254–265.

[210] S.-S. Guan and M. R. Luo, "A Colour-Difference Formula for Assessing Large Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 344–355, 1999.

[211] S.-S. Guan and M. R. Luo, "Investigation of Parametric Effects Using Large Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 356–368, 1999.

[212] H. Xu, H. Yaguchi, and S. Shioiri, "Estimation of Color-Difference Formulae at Color Discrimination Threshold Using CRT-generated Stimuli," *Optical Review*, vol. 8, no. 2, pp. 142–147, 2001.

[213] S. Shen and R. S. Berns, "Evaluating Color Difference Equation Performance Incorporating Visual Uncertainty," *Color Research and Application*, vol. 34, no. 5, pp. 375–390, 2009.

[214] R. S. Berns and B. X. Hou, "RIT-DuPont Supra-Threshold Color-tolerance Individual Color-difference Pair Dataset," *Color Research and Application*, vol. 35, no. 4, pp. 274–283, 2010.

[215] R. Shamey, L. M. Cárdenas, D. Hinks, *et al.*, "Comparison of Naïve and Expert Subjects on the Assessment of Small Color Differences," *Journal of the Optical Society of America A*, vol. 27, no. 6, pp. 1482–1489, 2010.

[216] J. Ma, H. S. Xu, M. R. Luo, *et al.*, "Color Appearance and Visual Measurements for Color Samples with Gloss Effect," *Chinese Optics Letters*, vol. 7, no. 9, pp. 869–872, 2009.

[217] S. G. Kandi and M. A. Tehrani, "Investigating the Effect of Texture on the Performance of Color Difference Formulae," *Color Research and Application*, vol. 35, no. 2, pp. 94–100, 2010.

[218] Z. N. Huang, H. S. Xu, M. R. Luo, *et al.*, "Assessing Total Differences for Effective Samples Having Variations in Color, Coarseness, and Glint," *Chinese Optics Letters*, vol. 8, no. 7, pp. 717–720, 2010.

[219] H. Wang, G. Cui, M. R. Luo, *et al.*, "Evaluation of Colour-Difference Formulae for Different Colour-Difference Magnitudes," *Color Research & Application*, vol. 37, no. 5, pp. 316–325, 2012.

[220] M. Melgosa, P. A. García, L. Gómez-Robledo, *et al.*, "Notes on the Application of the Standardized Residual Sum of Squares Index for the Assessment of Intra- and Inter-observer Variability in Color-Difference Experiments," *Journal of the Optical Society of America A*, vol. 28, no. 5, pp. 949–953, 2011.

[221] P. E. Shrout and J. L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.

[222] R. W. Pridmore and M. Melgosa, "Effect of Luminance of Samples on Color Discrimination Ellipses: Analysis and Prediction of Data," *Color Research and Application*, vol. 30, no. 3, pp. 186–197, 2005.

[223] W. R. J. Brown, "The Influence of Luminance Level on Visual Sensitivity to Color Differences," *Journal of the Optical Society of America*, vol. 41, no. 10, pp. 684–688, 1951.

[224] H. Xu, H. Yaguchi, and S. Shioiri, "Chromaticity-Discrimination Ellipses for Surface Colours," *Color Research and Application*, vol. 11, no. 1, pp. 25–42, 1986.

[225] M. Melgosa, M. M. Pérez, and A. El Moraghi, "Color Discrimination Results from a CRT Device: Influence of Luminance," *Color Research and Application*, vol. 24, no. 1, pp. 677–679, 1999.

[226] F. Carreño and J. M. Zoido, "The Influence of Luminance on Color-Difference Thresholds," *Color Research and Application*, vol. 26, no. 5, pp. 362–368, 2001.

[227] R. S. Berns, D. H. Alman, L. Reniff, *et al.*, "Visual Determination of Suprathreshold Color-Difference Tolerances Using Probit Analysis," *Color Research and Application*, vol. 16, no. 5, pp. 297–316, 1991.

[228] M. Melgosa, E. Hita, A. J. Poza, *et al.*, "Suprathreshold Color-Difference Ellipsoids for Surface Colors," *Color Research and Application*, vol. 22, no. 3, pp. 148–155, 1997.

[229] A. Yebra, J. A. García, J. L. Nieves, *et al.*, "Chromatic Discrimination in Relation to Luminance Level," *Color Research and Application*, vol. 26, no. 2, pp. 123–131, 2001.

[230] H. Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, 1931.

[231] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Harlow, United Kingdom: Pearson Education Limited, 2014.

[232] G. E. P. Box, "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, vol. 36, no. 3/4, pp. 317–346, 1949.

[233] H. C. Box, *Set Lighting Technician's Handbook: Film Lighting Equipment, Practice, and Electrical Distribution*. Burlington, MA, USA: Taylor & Francis, 2010.

[234] Commission Internationale de l'Éclairage, "Colorimetry – Part 6: CIEDE2000 Colour-difference Formula," *ISO/CIE 11664-6:2014*, 2014.

[235] H. W. Lilliefors, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[236] B. Rosner, "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, 1983.

[237] B. Iglewicz and D. C. Hoaglin, *How to Detect and Handle Outliers*. Milwaukee, WI: ASQC Quality Press, 1993.

[238] F. E. Grubbs, "Sample Criteria for Testing Outlying Observations," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 27–58, 1950.

[239] F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[240] A. N. Kolmogorov, "Sulla Determinazione Empirica di una Legge di Distribuzione," *Giornale dell'Istituto Italiano degli Attuari*, vol. 4, pp. 83–91, 1933.

[241] D. Öztuna, A. H. Elhan, and E. Tüccar, "Investigation of Four Different Normality Tests in Terms of Type 1 Error rate and Power under Different Distributions," *Turkish Journal of Medical Sciences*, vol. 36, no. 3, pp. 171–176, 2006.

[242] W. D Wayne, "Kolmogorov–Smirnov One-Sample Test," in *Applied Nonparametric Statistics*. Boston, MA, USA: PWS-Kent, 1990, pp. 319–330.

[243] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[244] S. S. Shapiro and R. S. Francia, "An Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215–216, 1972.

[245] N. M. Razali and B. W. Yap, "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[246] A. Ghasemi and S. Zahediasl, "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012.

[247] H.-Y. Kim, "Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis," *Restorative Dentistry and Endodontics*, vol. 38, no. 1, pp. 52–54, 2013.

[248] D. N. Joanes and C. A. Gill, "Comparing Measures of Sample Skewness and Kurtosis," *The Statistician*, vol. 47, no. 1, pp. 183–189, 1998.

[249] D. Cramer, *Basic Statistics for Social Research: Step-by-step Calculations and Computer Techniques Using Minitab*. Abingdon-on-Thames, United Kingdom: Routledge Taylor & Francis Group, 1997.

[250] J. F. Ratcliffe, "The Effect on the *t* Distribution of Non-normality in the Sampled Population," *Applied Statistics*, vol. 17, no. 1, pp. 42–48, 1968.

[251] S. S. Sawilowsky and R. C. Blair, "A More Realistic Look at the Robustness and Type II Error Properties of the *t* Test to Departures from Population Normality," *Psychological Bulletin*, vol. 111, no. 2, pp. 352–360, 1992.

[252] S. S. Sawilowsky and S. B. Hillman, "Power of the Independent Samples *t* Test under a Prevalent Psychometric Measure Distribution," *Journal of Consulting and Clinical Psychology*, vol. 60, no. 2, pp. 240–243, 1992.

[253] H. O. Posten, "Robustness of the Two-Sample *t* Test," in *Robustness of Statistical Methods and Nonparametric Statistics*, D. Rasch and M. L. Tiku, Eds. Dordrecht, Holland: D. Reidel Publishing Company, 1984, pp. 92–99.

[254] W. S. Gosset (Student), "The Probable Error of a Mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.

[255] E. Kreyszig, *Introductory Mathematical Statistics: Principles and Methods*. New York: John Wiley & Sons, 1970.

[256] D. G. Altman, *Practical Statistics for Medical Research*. Boca Raton, FL, USA: Chapman & Hall, 1991.

[257] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[258] H. B. Mann and D. R. Whitney, "On a Test of Whether one of two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[259] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd ed. New York: John Wiley & Sons, 2014.

[260] A. J. Vickers, "Parametric Versus Non-parametric Statistics in the Analysis of Randomized Trials with Non-normally Distributed Data," *BMC Medical Research Methodology*, vol. 5, no. 1, pp. 35–47, 2005.

[261] M. B. Brown and A. B. Forsythe, "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 364–367, 1974.

[262] H. Levene, "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, S. G. Ghurye, W. Hoeffding, *et al.*, Eds. Stanford, CA, USA: Stanford University Press, 1960, pp. 278–292.

[263] K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*. New York: John Wiley & Sons, 1965.

[264] W. U. Behrens, "Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen," *Landwirtschaftliche Jahrbücher*, vol. 68, pp. 807–837, 1929.

[265] R. A. Fisher, "The Fiducial Argument in Statistical Inference," *Annals of Human Genetics*, vol. 6, no. 4, pp. 391–398, 1935.

[266] R. A. Fisher, "The Asymptotic Approach to Behrens' Integral with further Tables for the *d* Test of Significance," *Annals of Human Genetics*, vol. 11, no. 1, pp. 141–172, 1941.

[267] T. P. Hettmansperger and J. W. McKean, *Robust Nonparametric Statistical Methods*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2011.

[268] B. L. Welch, "The Significance of the Difference Between Two Means when the Population Variances are Unequal," *Biometrika*, vol. 29, no. 3/4, pp. 350–362, 1938.

[269] B. L. Welch, "The Generalization of "student's"problem when Several Different Population Variances are Involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.

[270] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. San Diego, CA, USA: GraphPad Software, Inc, 2003.

[271] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, *et al.*, "Modeling of the Bacterial Growth Curve," *Applied and Environmental Microbiology*, vol. 56, no. 6, pp. 1875–1881, 1990.

[272] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

[273] R. M. Warner, *Applied Statistics: From Bivariate Through Multivariate Techniques*, 2nd. London, United Kingdom: SAGE Publications, Ltd, 2013.

[274] J. E. Pustejovsky, "Converting from d to r to z when the Design uses Extreme Groups, Dichotomization, or Experimental Control," *Psychol. Methods*, vol. 19, no. 1, pp. 92–112, 2014.

[275] M. Borenstein, "Effect Sizes for Continuous Data," in *The Handbook of Research Synthesis and Meta-Analysis*, H. Cooper, L. V. Hedges, and J. C. Valentine, Eds., 2nd. New York, NY, USA: Sage Foundation, 2009, pp. 221–235.

[276] A. Field, *Discovering Statistics using IBM SPSS Statistics*, 4th. London, United Kingdom: SAGE Publications, Inc., 2013.

[277] J. Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis using IBM SPSS*, 6th. Philadelphia, PA, USA: Mcgraw-Hill Higher Education, Open University Press, 2016.

[278] R. R., "Parametric measures of effect size," in *The Handbook of Research Synthesis*, H. Cooper and L. V. Hedges, Eds. New York, NY, USA: Russel SAGE Foundation, 1994, pp. 231–244.

[279] J. P. Stevens, *Applied Multivariate Statistics for the Social Science*, 5th. New York, NY, USA: Taylor & Francis Group, 2009.

[280] I. T Hernádvölgyi, *Generating Random Numbers from the Multivariate Normal Distribution*, `ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia013_2s09/material_de_apoio/gen_rand_multivar.pdf`, [Online; accessed 14-March-2018], 1998.

Part 5

MEMORY COLORS AND THE ASSESSMENT OF COLOR QUALITY

"The reflection is for the colors what the echo is for the sound"

– Joseph Joubert [97]

5 MEMORY COLORS AND THE ASSESSMENT OF COLOR QUALITY

Being based on the findings of the previous chapters, an updated memory-based color quality metric should be derived which evaluates the degree of similarity between the color appearance of certain familiar test objects rendered by the test light source under inspection and their respective memory color representations. Here, the degree of similarity will be assessed by using a set of similarity distribution functions fitted to the results of the previously conducted color appearance rating experiments. A meta-correlation analysis should subsequently be performed to investigate the predictive performance of the updated metric for the color rendition evaluation of white light sources with respect to the color perception attributes of visual appreciation and color preference.

5.1 DEFINITION OF A MEMORY COLOR PREFERENCE INDEX

In order to assess the color quality of conventional and solid-state light sources with regard to visual appreciation, several different color quality metrics have been proposed in the literature among which Smet *et al.*'s MCRI [23, 52–54] is the only one that is based on the evaluation of memory colors. Apart from making use of an arguable test object selection and color space (see Sec. 3.2.4), Smet *et al.* provided an excellent methodology and a very intuitive approach. Consequently, their principal definitions should be adopted here with some additional improvements regarding the set of familiar test objects, their representation in a perceptually uniform color space, and the inclusion of further impact factors such as the adapted white point and the cultural background of the observers.

For this purpose, the color appearance of a set of twelve familiar real objects was investigated in a series of visual rating experiments whose results were reported in the previous chapter. The familiar test objects were selected based on the outcome of an online survey which was intended to find for each hue region the most frequently associated colored object among all participants, which is supposed to be a good approximation of a typical memory color according to the original definition of Hering [32]. The following twelve test objects were eventually chosen: Asian skin, banana, blueberry, blue jeans, broccoli, butternut squash, carrot, Caucasian skin, concrete flowerpot, green salad, red cabbage, red rose. Each of these test objects was presented to a group of 15 German observers and, in a follow-up experiment, to a group of 15 Chinese observers while being illuminated by a combination of a four-channel LED light source setting the ambient adaptation conditions and an LCD projector used for shifting the color appearance of the respective object resulting in a total number of approximately 65 color variations per test object. In addition, two different adaptation conditions at 3200 K and 5600 K ambient illumination were examined so that for each test object four different experimental runs were accomplished, i.e., German observers at 3200 K, German observers at 5600 K, Chinese observers at 3200 K, and Chinese observers at 5600 K.

In each case, the observers were asked to rate the color appearance of the currently presented test object according to their preference of how they thought the respective object should look like in reality on a semi-semantic five-level scale. The mean color appearance ratings for each familiar test object i were subsequently modeled in CIECAM02-UCS chromaticity space by utilizing a bivariate Gaussian function defined in Eqs. 4.15 and 4.16. This

Table 5.1 – Summary of the relevant fit parameters describing the similarity distribution functions of the twelve familiar test objects assessed by the Chinese and German observers at 3200 K ambient illumination. Parameters a_3 and a_4 give the locations of the centroids of the distribution functions which are defined to be the most likely representations of the objects' memory colors in CIECAM02-UCS chromaticity space. The size, shape, and orientation of the similarity distribution functions are determined by the parameters a_5 to a_7.

	Chinese observers at 3200 K adapted white point					German observers at 3200 K adapted white point				
Test object	a_3	a_4	a_5	a_6	a_7	a_3	a_4	a_5	a_6	a_7
Asian Skin	12.9180	10.1611	0.0778	0.0524	−0.0096	13.0361	10.0919	0.0870	0.0507	−0.0094
Banana	3.8769	28.2924	0.0199	0.0075	0.0043	5.0739	31.7735	0.0206	0.0054	0.0035
Blueberry	−2.8453	−12.8145	0.0409	0.0127	−0.0032	−3.5445	−12.6203	0.0535	0.0166	−0.0044
Blue Jeans	−7.6266	−17.1651	0.0357	0.0107	−0.0015	−8.1133	−19.7052	0.0455	0.0126	−0.0036
Broccoli	−9.6176	14.3775	0.0093	0.0013	0.0005	−8.6225	9.2019	0.0131	0.0060	−0.0004
Butternut Squash	15.7546	20.6420	0.0160	0.0116	−0.0011	14.7260	20.4684	0.0265	0.0201	−0.0032
Carrot	27.4901	26.6343	0.0100	0.0149	−0.0003	25.0920	25.0149	0.0136	0.0127	−0.0009
Caucasian Skin	12.7231	9.4047	0.0788	0.0539	−0.0060	12.1936	9.2285	0.0879	0.0563	−0.0038
Concrete Flowerpot	−0.8219	−3.5351	0.0200	0.0087	−0.0054	−0.8031	−3.3339	0.0369	0.0150	−0.0057
Green Salad	−13.5223	23.1321	0.0313	0.0114	−0.0012	−13.5855	24.1868	0.0240	0.0193	0.0004
Red Cabbage	11.9059	−16.9319	0.0186	0.0092	−0.0061	12.8639	−13.6538	0.0360	0.0137	−0.0067
Red Rose	37.1120	13.2141	0.0440	0.0099	0.0043	36.6615	12.7504	0.0259	0.0278	−0.0064

model basically requires seven parameters $a_{i,1}$ to $a_{i,7}$. In contrast to the parameters $a_{i,3}$ to $a_{i,7}$ which define the conceptually important similarity distribution function $S_i(\mathbf{x}_i)$ of the i^{th} test object, the parameters $a_{i,1}$ and $a_{i,2}$ are simply used to adjust the Gaussian to the applied five-level rating scale. With these two parameters being therefore superfluous for the evaluation of the degree of similarity between the object's color appearance and its respective memory color representation, only the test objects' similarity distribution functions should be considered in the following. The functions' centroids $\boldsymbol{\mu}_i$, giving the most likely representations of the test objects' memory colors, are located at $(a_{i,3}, a_{i,4})^\mathsf{T}$, whereas the size, shape, and orientation of the similarity distributions are determined by the inverse of the respective covariance matrix $\boldsymbol{\Sigma}_i$ which is defined by the parameters $a_{i,5}$ to $a_{i,7}$. To get a better overview, Tables 5.1 and 5.2 summarize these sets of relevant fitting parameters for all twelve familiar test objects assessed by both cultural observer groups which were adapted to the ambient illumination at 3200 K and 5600 K, respectively.

With the fitting parameters of the twelve familiar objects being known, the assessment of the color quality of a certain test light source can eventually be based on the evaluation of the degree of similarity between the color appearance of these objects rendered by the test light source and their respective memory color representations. For this purpose, the CIECAM02-UCS chromaticities $\mathbf{x}_i = \left(a'_{\mathrm{M},i}, b'_{\mathrm{M},i}\right)^\mathsf{T}$ of each familiar test object i as perceived under the test light source are calculated first by using the measured spectral reflectances of the test objects exemplarily illustrated in Fig. 4.14 and the CIE standard 10° observer. The SPD of the test light source must be known or measured by a spectroradiometer. Next, the function values of the corresponding similarity distributions $S_i(\mathbf{x}_i)$ are determined, where

$$\begin{aligned} S_i(\mathbf{x}_i) &= \exp\left(-\frac{1}{2}\left((\mathbf{x}_i - \boldsymbol{\mu}_i)^\mathsf{T}\,\boldsymbol{\Sigma}_i^{-1}\,(\mathbf{x}_i - \boldsymbol{\mu}_i)\right)\right) \\ &= \exp\left(-\frac{1}{2}\left((\mathbf{x}_i - \boldsymbol{\mu}_i)^\mathsf{T}\begin{pmatrix} a_{i,5} & a_{i,7} \\ a_{i,7} & a_{i,6}\end{pmatrix}(\mathbf{x}_i - \boldsymbol{\mu}_i)\right)\right) \end{aligned} \tag{5.1}$$

describes the degree of similarity with each object's memory color, i.e., the closer the function value is to one, the better the agreement. Rescaling these individual similarity indicators S_i to a 0–100 range gives the specific memory color preference indices $R_{\mathrm{MCPI},i}$ defined by

$$R_{\mathrm{MCPI},i} = 100 \cdot S_i. \tag{5.2}$$

Bearing in mind that for the visual appreciation of a perceived lighting scene certain colors were shown to be more important than others [20, 195, 196], additional weighting factors for the specific indices should be included here. Hence, the final general memory color preference index R_{MCPI} is obtained by calculating the weighted geometric mean of the twelve individual $R_{\mathrm{MCPI},i}$ values, which reads

$$R_{\mathrm{MCPI}} = \prod_{i=1}^{12} \left(R_{\mathrm{MCPI},i}\right)^{p_i}, \tag{5.3}$$

where p_i with $i = 1, ..., 12$ denotes the individual weighting factors whose appropriate values will be determined from meta-correlation analysis (see Sec. 5.2.3). According to Smet *et al.* [23], the geometric mean should be preferred over the arithmetic mean in the current application since as stated by these authors it is i) less susceptible to outliers, which in general makes the color quality metric more robust, and ii) it better suits the exponential nature of the function values of the similarity distributions.

So far, the calculation scheme of the new memory color preference index R_{MCPI} is quite similar to the one provided for Smet's MCRI R_{m} [23, 52–54] – except for the additionally introduced weighting factors. However, in order to account for the findings of the previous chapter, some further modifications were necessary. First, a decision algorithm based on CCT calculation was implemented to approximate the impact of the white point of the test light source under investigation on the memory color assessments. If the corresponding CCT is smaller than 4000 K, the test light source is considered to be rather warm white which implies the selection of the parameters of Table 5.1 to define the test objects' similarity distribution

Table 5.2 – Summary of the relevant fit parameters describing the similarity distribution functions of the twelve familiar test objects assessed by the Chinese and German observers at 5600 K ambient illumination. Parameters a_3 and a_4 give the locations of the centroids of the distribution functions which are defined to be the most likely representations of the objects' memory colors in CIECAM02-UCS chromaticity space. The size, shape, and orientation of the similarity distribution functions are determined by the parameters a_5 to a_7.

	Chinese observers at 5600 K adapted white point					German observers at 5600 K adapted white point				
Test object	a_3	a_4	a_5	a_6	a_7	a_3	a_4	a_5	a_6	a_7
Asian Skin	10.2878	10.8645	0.0615	0.0465	−0.0101	10.2923	10.3590	0.0627	0.0336	−0.0050
Banana	4.4418	30.8273	0.0195	0.0121	0.0029	5.0830	32.5946	0.0237	0.0126	0.0052
Blueberry	−2.4215	−10.5405	0.0351	0.0153	−0.0046	−3.3163	−12.2902	0.0365	0.0161	−0.0044
Blue Jeans	−7.1205	−14.7482	0.0294	0.0122	−0.0035	−7.7858	−17.3582	0.0486	0.0134	−0.0053
Broccoli	−8.3038	16.1013	0.0419	0.0167	0.0025	−8.2703	15.3796	0.0367	0.0391	0.0043
Butternut Squash	14.1887	19.6973	0.0143	0.0138	−0.0018	15.3767	20.5664	0.0205	0.0180	−0.0033
Carrot	24.9891	24.6469	0.0230	0.0183	0.0022	25.6425	25.6319	0.0198	0.0154	−0.0034
Caucasian Skin	8.4641	8.2530	0.0481	0.0302	−0.0060	9.7097	8.7088	0.0620	0.0362	−0.0064
Concrete Flowerpot	−0.4814	−1.3515	0.0289	0.0109	−0.0040	0.0840	−1.1743	0.0395	0.0162	−0.0048
Green Salad	−9.5558	26.0556	0.0312	0.0257	0.0067	−10.4320	26.6978	0.0288	0.0262	0.0029
Red Cabbage	10.4067	−14.6458	0.0262	0.0128	−0.0057	10.5638	−14.0777	0.0262	0.0146	−0.0080
Red Rose	34.8656	11.2483	0.0724	0.0206	0.0066	34.6755	11.1227	0.0719	0.0353	−0.0049

Table 5.3 – Summary of the relevant fit parameters describing the similarity distribution functions of the twelve familiar test objects assessed by an assumed global average observer at both adaptation conditions. Parameters a_3 and a_4 give the locations of the centroids of the distribution functions which are defined to be the most likely representations of the objects' memory colors in CIECAM02-UCS chromaticity space. The size, shape, and orientation of the similarity distribution functions are determined by the parameters a_5 to a_7.

	Global observer at 3200 K adapted white point					Global observer at 5600 K adapted white point				
Test object	a_3	a_4	a_5	a_6	a_7	a_3	a_4	a_5	a_6	a_7
Asian Skin	12.9776	10.1264	0.0822	0.0515	−0.0095	10.2899	10.6439	0.0621	0.0398	−0.0075
Banana	4.5023	29.7617	0.0197	0.0063	0.0038	4.8156	31.7194	0.0214	0.0120	0.0038
Blueberry	−3.2205	−12.7014	0.0461	0.0144	−0.0037	−2.9071	−11.5154	0.0389	0.0171	−0.0052
Blue Jeans	−7.8787	−18.4912	0.0404	0.0115	−0.0027	−7.7717	−15.6493	0.0360	0.0116	−0.0048
Broccoli	−8.8949	10.8201	0.0124	0.0032	0.0000	−8.2913	15.6197	0.0406	0.0270	0.0037
Butternut Squash	15.1462	20.5524	0.0208	0.0153	−0.0018	14.8458	20.1952	0.0173	0.0157	−0.0025
Carrot	26.1198	25.9112	0.0113	0.0127	−0.0007	25.2292	25.0584	0.0211	0.0165	−0.0005
Caucasian Skin	12.4669	9.3178	0.0828	0.0551	−0.0052	9.0753	8.4607	0.0522	0.0320	−0.0062
Concrete Flowerpot	−0.8350	−3.4252	0.0266	0.0115	−0.0056	−0.1393	−1.1642	0.0324	0.0130	−0.0043
Green Salad	−13.5297	23.7977	0.0286	0.0148	−0.0004	−9.9728	26.3866	0.0305	0.0261	0.0052
Red Cabbage	12.5222	−14.9915	0.0261	0.0113	−0.0066	10.4791	−14.3666	0.0264	0.0138	−0.0069
Red Rose	36.9028	12.8961	0.0362	0.0174	0.0007	34.7823	11.1809	0.0717	0.0279	0.0008

functions. If, on the other hand, the calculated CCT is greater or equal 4000 K, the test light source is assumed to be rather cool white and the parameters given in Table 5.2 are chosen. Second, two different cultural-specific MCPIs should be created, one Chinese and one German version, in order to account for the inter-cultural differences in the evaluation of the color quality of light sources between these two cultural observer groups. Hence, by choosing the appropriate fit parameters from Tables 5.1 and 5.2, i.e., the ones that are assigned to either the Chinese or the German observer group for a specific adapted white point, both the impact of the observers' cultural background as well as the effect of different adaptation conditions can be considered for the memory-based evaluation of color rendition.

Furthermore, a global average observer was defined by pooling and averaging for each test object and adaptation condition the rating data of both cultural observer groups. The obtained mean observer ratings were subsequently modeled by applying again a bivariate Gaussian fitting, where the corresponding parameters are summarized in Table 5.3. Basically, this global average observer should serve as a reference to investigate whether a more universally valid MCPI can be proposed or if the impact of the observed inter-cultural variations on the color appearance ratings of familiar objects reported in Sec. 4.4.3 also has a significant effect on the predictive performance of the MCPI metric. For this purpose, a meta-analysis of the Spearman correlation coefficients [281, 282] between the metric predictions for various light sources of several psychophysical studies and the corresponding observer ratings will be performed in the following section.

5.2 PERFORMANCE VALIDATION BASED ON META-CORRELATION ANALYSIS

In this section, the experimental data of several psychophysical studies providing a wide range of different lighting scenarios were collected in order to eventually perform a meta-correlation analysis to evaluate and compare the predictive power of the MCPI and various other color quality metrics with respect to the concept of visual appreciation. The experimental data contain the SPDs of the investigated light sources together with the corresponding

observer preference ratings. Based on their principle study design, two different kinds of experiments should be considered: The first group made use of metameric lighting scenarios, i.e., all tested light sources to be assessed and compared within a single experimental trial showed the same CCT but different spectral characteristics. The second group, on the other hand, applied multi-CCT scenarios, which means that not only the emitted spectra but also the corresponding CCTs of the light sources are varied for the comparison performed by the observers.

5.2.1 *Overview of the Collected Studies*

The subsequent paragraphs give an overview of the studies being used for running the meta-correlation analysis. For further details, the interested reader is referred to the original publications. Please note that in most cases, only the average observer ratings as reported therein were accessible, while the corresponding SPDs of the light sources – if not available otherwise (e.g., from personal communication with the respective authors) – were digitized using the powerful WebPlotDigitizer tool written by Ankit Rohatgi [283].

WEI *et al.* [284] (METAMERIC LIGHTING, 2014) In this study, a side-by-side comparison of two identically furnished rooms exhibiting the same colored objects and still life arrangements should be performed. Each room was illuminated by a 3000 K light source giving an illuminance level at the objects' location of approximately 250 lx. In one of the rooms, the illumination was realized by standard blue-pumped LED A19 lamps, whereas in the other room similar LED lamps but with diminished yellow emission were used. A paired-comparison experiment was conducted, where 87 participants compared and rated the color appearance of the test objects under both illuminations according to their preference using a six-point rating scale.

JOST-BOISSARD *et al.* [154, 195] (METAMERIC LIGHTING, 2009 AND 2014) In two series of visual experiments, several aspects of perceived color quality of LED light sources were examined. In the first series, a total number of 40 subjects were invited to assess the color rendering of four different arrangements of fruits and vegetables illuminated by LED light sources in comparison to some reference illumination at two different CCTs of 3050 K and 3950 K in terms of attractiveness, naturalness, and suitability. For this purpose, a viewing cabinet with two identical compartments (actually three, but only two of them were used for the experiment) was equipped with the reference illuminant in one of the compartments and with six different LED clusters in the other one. For the experiments at 3050 K, a halogen light source was used as the reference, whereas in case of the higher CCT a fluorescent light source was chosen. The corresponding illuminance levels were 230 lx and 220 lx, respectively. For each of these cases, the relative intensities of the LED clusters were adjusted to match the illuminance level and white point of the reference light source resulting in six different test spectra for each CCT with minimal differences regarding the white point of adaptation. With the subsequently performed experiments adopting a forced-choice paired-comparison method, the participants were asked to choose for each presented arrangement of test objects the appropriate lighting condition that best suited their subjective judgments. For the following meta-correlation analysis, only the data regarding attractiveness will be used.

In the second series of experiments, a quite similar investigation using the same tripartite viewing cabinet was conducted for various light sources exhibiting CCTs of 3000 K and 4000 K. The corresponding illuminance levels were 230 lx and 210 lx, respectively. This time,

no fixed reference was given and the forced-choice paired-comparison was performed for each possible combination of light sources with respect to the color quality aspects of naturalness, colorfulness, visual appreciation, and fidelity (color difference). At 3000 K, two conventional (halogen and fluorescent) and seven LED light sources were assessed by 45 observers, while one conventional (fluorescent) and seven LED light sources were assessed by 36 observers in the 4000 K case. For each CCT, two different scenarios were presented: i) a plate of colored fruits and vegetables and ii) a Macbeth ColorChecker® chart. Only the data for visual appreciation will be used in the following.

SZABÓ *et al.* [285] (METAMERIC LIGHTING, 2016) Two different home lighting scenarios, i.e., kitchen and living room, were investigated in terms of the perceived color quality aspects pleasantness, vividness, and naturalness. A total number of 69 young and 28 elderly observers were invited to participate in the experiments. For the kitchen and living room scenario LED luminaires with CCTs of 4000 K and 3000 K were chosen, respectively. These LED luminaires consisted of a mix of 20 colored and phosphor-converted LEDs ensuring high flexibility in creating various light sources with different SPDs but the same CCT. Thus, for each lighting scenario ten different SPDs were optimized, of which five showed a constant and the other five an inconstant FCI value clustering the test light sources into two different groups. In all cases, the illuminance level was set to 350 lx. For each scenario and light source group, a paired-comparison experiment was conducted, in which the participants were asked to assess the color quality of the presented scene on a five-level rating scale. In order to be consistent with the other studies used in the meta-correlation analysis, only the rating data of the young observers will be considered. In addition, the results for constant FCI are excluded because these are special cases that would from the beginning adulterate the comparison of the predictive performance of different metrics in color preference evaluation (i.e., zero correlation would be found between the FCI values and the preference ratings).

SMET *et al.* [23] (METAMERIC LIGHTING, 2010) A paired comparison experiment in a bipartite viewing booth was performed to determine the perceived color quality of six different 2700 K conventional and solid-state light sources. A total number of 92 observers took part in the experiment. They were shown, in successive order, an arrangement of colored objects illuminated by a pair of different light sources. For each possible pair, the observers were asked to rate the difference in color quality on a seven-point rating scale regarding the aspects of preference, fidelity, vividness, and attractiveness, where, in order to be consistent with the other studies, only the results for the latter were extracted to be used in the meta-analysis. In all cases, the illuminance level at the objects' location was approximately 250 lx.

VANRIE [286] (METAMERIC LIGHTING, 2009) A panel of 30 subjects took part in a visual experiment where the task was to rank a mixture of six different conventional and LED light sources in the order of their perceived color quality with regard to the aspects of attractiveness and naturalness. The six luminaires all showed a CCT of approximately 3000 K. Identical arrangements of single colored objects (a red tomato, a green apple, and a yellow lemon) were placed inside six different viewing cabinets and illuminated by one of the six test light sources causing in each case an illuminance level of 310 lx measured at the position of the objects. The six viewing cabinets were simultaneously presented to the observers. During the experiment, the order of the light sources was switched using a latin-square method [287] in order to avoid possible bias effects. For each light source, the evaluation of the per-

ceived color quality was given by the average observer rankings, of which only the results for attractiveness will be considered in the following.

IMAI *et al.* [288, 289] (METAMERIC LIGHTING, 2012 AND 2013) In a series of two similar experiments, Imai *et al.* investigated the relation between color preference evaluation and the saturation of object colors for LED illumination. For both experiments, a specifically designed experimental booth simulating a real room was used. A spectrally adjustable lighting system was mounted into the ceiling of the booth consisting of 1500 high-power LEDs of eight different color channels. In the first experiment, various light spectra at three different CCTs were examined. At 3000 K, 5000 K, and 6500 K data of eight, two, and six different SPDs were available, respectively. Thirty subjects took part in the experiment and were asked to evaluate for each lighting situation an arrangement of multicolored object samples, which had been placed on a desk inside the experimental booth, according to their personal appreciation adopting a bipolar seven-level rating scale.

In the second experiment, more or less the same setup and rating method were applied. The only difference was that instead of assessing an arrangement of several colored objects, the observers rated only one object at a time. Color Preference data of ten different SPDs with a CCT of 5000 K were available for the test objects of red apple, orange, and cyan yarn. Again, ratings of thirty subjects were collected. In both experiments, the illuminance level at the test objects' location was set to 700 lx.

JOST *et al.* [290] (METAMERIC LIGHTING, 2013) An experiment was performed to investigate the perceived color rendering of face complexion and hair under LED and fluorescent light sources. For this purpose, a viewing booth was designed that enabled the evaluation of the subjects' own appearance illuminated by 20 different light sources which all showed a CCT of 2700 K. In order to provide a realistic lighting scenario as it could be imagined to be present in the homes of the participants, the booth was equipped with a mirror which was edged on his left and right hand side by the test light sources hidden under diffusing plates. In such a way, an illuminance level of 220 lx could be achieve on the subjects' face when seated in front of the mirror inside the viewing booth. A total number of 63 observers which differed in their skin and/or hair tones participated in the experiment. The 20 light sources were presented in random order. For each of these settings, the observers were asked to rate their visual appreciation of the color rendering on a categorical seven-point rating scale. These ratings should be performed for their skin and hair (observed through the mirror) as well as for a Macbeth ColorChecker® chart which was introduced as a third target. All of these collected preference ratings will be considered in the following.

TSUKITANI [291, 292] (METAMERIC LIGHTING, 2013 AND 2016) Another series of paired-comparison experiments was conducted to evaluate the color quality of various light sources illuminating identical natural objects which were again placed side-by-side in a bipartite viewing cabinet. In the first experiment, a green plant as well as red and yellow leaves were used as test objects. Two different CCTs at 2800 K and 3000 K were adopted. In both cases, the illumination on the reference side of the viewing cabinet was realized by commercially available LED lamps exhibiting the respective CCT. Regarding the evaluation compartment of the viewing cabinet, a spectrally tunable multi-channel LED light source was installed. Based on this device, a total number of 14 different SPDs with a CCT of 2800 K and 15 different SPDs with a CCT of 3000 K were optimized. In all cases, the illuminance at the test objects' position was 150 lx, which equaled the value measured in the reference

compartment. During the experiment, the test illuminants were compared with the reference having the same CCT. For each case, observers were asked to rate the perceived color quality in terms of their visual appreciation by two different methods. The first method was a magnitude estimation, i.e., observers should indicate by which factor the attractiveness of the scene illuminated by the test light source increased in comparison to the reference illumination. For the second method, the assessment of color quality and visual appreciation was performed by five different bipolar word pairs marked on a corresponding seven-point rating scale. For the subsequent meta-correlation analysis only data of the former method were accessible.

In the second experiment, a similar experimental setup was used. This time, various fruits, vegetables, and other groceries as well as some flowers were used as test objects which were individually presented to the observers and illuminated by ten different SPDs, five showing a CCT of 3000 K and the other five a CCT of 5000 K, causing an illuminance level of approximately 500 lx at the bottom of the viewing cabinet. In total, 70 observers took part in the experiment whose task was to evaluate the color appearance of the presented test object for each lighting condition in terms of vividness, naturalness, preference, and deliciousness on a semi-semantic nine-level scale. For each test spectrum, this evaluation was performed twice by adopting both a relative and an absolute rating method. While for the latter no reference was given (only one compartment of the viewing cabinet was lit), the former requires the observers to compare the test object's color appearance under the test light source with some reference illumination of the same CCT and illuminance provided in the reference compartment. From the publication the mean ratings for the various color quality aspects averaged over all test objects were available, of which only the ones related to preference are used in the following. Furthermore, the relative rating results are neglected for the meta-correlation analysis since a good memory-based color quality metric should accurately predict the observer preference even in the absence of an explicit reference.

HOUSER *et al.* [293] (METAMERIC LIGHTING AND MULTI-CCT, 2005) Eighteen expert and twenty-three naïve observers were invited to participate in a pilot study to examine the color quality of four different prototype fluorescent lamps in terms of preference, naturalness, and colorfulness. Two of these lamps had a CCT of 3500 K and the other two of 6500 K. A forced-choice paired comparison experiment was conducted on two identically furnished real-size office rooms presented side-by-side to the observers. All possible combinations were shown and for each pair of light sources, creating both metameric and multi-CCT rating scenarios, observers had to answer 21 forced-choice questions. In all cases, the illuminance level was set to 538 lx measured on the office desks. The key data of the questionnaire were summarized by the authors regarding the above mentioned aspects of color quality. With expert and naïve observers showing similar responses and the same general trend, the reported data of these two observer groups were averaged, of which only the preference results will be used for the meta-correlation analysis of the current work.

HUANG *et al.* [294] (MULTI-CCT, 2017) In this work, a series of two different psychophysical experiments was conducted. The first one was again performed using a viewing cabinet consisting of a single compartment equipped with several commercially available and color-tunable LED light bulbs. During the experiment, various test objects were examined, which included four groups of fruit and vegetable arrangements similar to the work of Jost-Boissard *et al.* [195], five traditionally Chinese calligraphies written on papers of different colors, four pieces of artwork of different size and color properties, and a bunch of multicolored flowers. Making use of the color-tunability feature of the light bulbs, several different SPDs with uni-

formly sampled CCTs ranging from 2500 K to 6500 K exhibiting an illuminance level of 200 lx at the test objects' position were optimized. The number of observers assessing each lighting scenario ranged from 36 to 60 which were asked to rate the light quality based on their personal preference using in most cases a seven-point rating scale – only for the evaluation of the artwork scenarios a five-level rank ordering method was chosen.

In the second experiment, a wall-painting should be assessed. The experiment took place in the museum where it is usually exhibited. For collecting observer preference data, the same optimized SPDs were used as for the rating of the color appearance of the artworks presented in the viewing cabinet, with the sole difference that the light source was dimmed in such a way that the vertical illuminance measured in the center of the wall-painting was 50 lx for all CCTs. Twenty observers participated in the experiment with the color preference rating being performed on the same seven-point rating scale as before.

NARENDRAN *et al.* [295] (MULTI-CCT, 2002) Two different experiments were conducted to investigate the color preference of human observers in the context of LED-based reading lights. At first, a paired comparison experiment was conducted using a bipartite viewing cabinet where one side was equipped with two reference light sources (halogen and RGB LED with low CRI) while the other housed six different test light sources. For all combinations of reference and test light source, observers were asked to indicate in each case their preference for the lighting in the left or right compartment. An equal preference rating was also possible. In the second part of the experiment, only one of the compartments was lit with observers being invited to express their preference for the respective lighting on a seven-level rating scale. In total, 30 subjects took part in the experiment. The light sources all showed approximately the same illuminance level of 200 lx. With no significant differences being found between both evaluation methods, only the data of the seven-level rating experiment are used for the following meta-correlation analysis.

WANG *et al.* [296] (MULTI-CCT, 2017) Last but not least, the results of a questionnaire-based forced-choice experiment performed by Wang *et al.*, investigating the impact of different combinations of CCT and illuminance on human preference and visual comfort ratings, shall be considered for the subsequent meta-correlation analysis. During the experiment, a printed photograph containing four typical images of natural objects (e.g., blue sky, green grass, human skin, etc.) was placed inside a multi-channel LED lighting booth and illuminated by specifically optimized SPDs exhibiting various CCTs and illuminance levels. In total, twelve different CCTs ranging from 2000 K to 100 000 K at three different illuminance levels (350 lx, 500 lx, and 1000 lx) were involved. For each combination, a preference and comfort estimation was performed based on the forced-choice answers to the provided questionnaire given by twelve different observers. The questions concerned the two possible scenarios of working and relaxing, of which only the results of the latter were chosen to be used in the following. As stated by Liu *at al.* [297], a working scenario would be more related to performance rather than to color preference. Furthermore, with CCT values higher than 10 000 K being very rare in usual lighting applications, the preference data for CCTs larger than that were omitted for the current study. Please note that only results averaged over all three illuminance conditions were available from the original publication. However, as stated by the authors, the preference estimations were quite consistent among the different illuminance levels and, therefore, averaging the observer ratings is considered to be no problem for the intended meta-correlation analysis.

Finally, it should be noted that there are significantly more studies investigating metameric lighting scenarios than those comparing multi-CCT conditions regarding the color preference evaluation of human observers. Nevertheless, the corresponding four publications listed above are considered to be sufficiently comprehensive to give a good indication for the general predictive performance of color quality metrics being applied to such lighting scenarios.

5.2.2 *Conceptional Design of the Meta-correlation Analysis*

Based on the compilation of quite a large number of studies dealing with the color quality aspects of visual appreciation and preference presented in the previous section, a meta-analysis of the Spearman correlation coefficients between the predictions of a certain color quality metric for various light source and the corresponding observer preference ratings should be performed. By definition, the calculation of a Spearman correlation does not make any assumptions about the nature of the underlying data and, basically, provides a measure for the metrics' ability to correctly rank light sources regarding a specific color quality aspect. The method for running a meta-correlation analysis adopted in this thesis was originally proposed by Hunter and Schmidt [298]. Since then it has successfully been applied in various fields of research when correlations should be analyzed over a wide range of various studies being addressed to a similar topic but using quite different methods for data acquisition and evaluation.

In principle, the method of Hunter-Schmidt (HS) tries to estimate the true correlation between two variables by calculating a weighted average correlation corrected for additional study artifacts which in general attenuate the actual correlation and typically lead to an underestimation of the true variance. Here, the basic sampling error corrected, weighted average correlation coefficient $\bar{r}$ is given by

$$\bar{r} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i}, \tag{5.4}$$

where r_i represents the individual Spearman correlation coefficient and N_i the respective observer number of the i^{th} lighting scenario with $i = 1, ..., k$, where k gives the total number of lighting scenarios extracted from the bunch of psychophysical studies discussed in the previous section. This simple correction for sampling error is also referred to a so-called bare-bones meta-correlation analysis. However, as stated in the textbook of Hunter and Schmidt [298], additional artifacts and imperfections are present in every study design which basically cause a non-negligible attenuation of the true correlation. Fortunately, they also provided a couple of methods allowing for the correction of some of these artifacts. Based on the information that could be extracted from the previous studies, the following errors and artifacts can be corrected for: i) bare-bones sampling error, ii) heterogeneity between different studies, iii) restriction or enhancement of range, iv) attenuation due to idiosyncrasy in the perception of the observers called halo (e.g., inter-observer variability), and v) statistical bias in the sample correlation.

Since the basic sampling error due to variations in sample size between the different studies is automatically corrected by applying Eq. 5.4, the following discussion should focus on the remaining four artifacts. Regarding for example the correction of study heterogeneity, an unbiased estimator $\hat{\bar{r}}$ of the true correlation can be obtained by adjusting the study weights in Eq. 5.4 from N_i to the true or optimal weights given by [299]

$$N_i^{\text{opt.}} = \frac{1}{\hat{\tau}^2 + N_i^{-1}}, \tag{5.5}$$

where $\hat{\tau}^2$ is the so-called heterogeneity estimator. Several different of such estimators can be found in the literature [300, 301]. The one that should be adopted here was originally defined by Hunter and Schmidt [298] and reads

$$\hat{\tau}^2_{\mathrm{HS}} = \frac{Q-k}{\sum_{i=1}^{k} N_i}, \tag{5.6}$$

where

$$Q = \sum_{i=1}^{k} N_i \left(r_i - \bar{r}\right)^2, \tag{5.7}$$

so that Eq. 5.4 changes to

$$\hat{\bar{r}} = \frac{\sum_{i=1}^{k} N_i^{\mathrm{opt.}} r_i}{\sum_{i=1}^{k} N_i^{\mathrm{opt.}}}. \tag{5.8}$$

Please note that in cases where $Q < k$, the heterogeneity estimator $\hat{\tau}^2_{\mathrm{HS}}$ would be negative and, therefore, must be truncated to zero.

The next source of attenuation that should be addressed is the range restriction due to study design. As stated by Hunter and Schmidt [298], correlations across different studies are only directly comparable if they are computed on samples from populations with the same standard deviation on the independent variable which in the present case is the respective metric score. Hence, in order to eliminate the effect of range variations between different studies from the meta-correlation analysis, a range correction formula is used which projects all correlations onto the same reference standard deviation σ_0. This correction for restriction in range for the i^{th} lighting scenario is given by

$$r_{0,i} = \frac{U_x r_i}{\sqrt{\left(U_x^2 - 1\right) r_i^2 + 1}}, \tag{5.9}$$

where

$$U_x = \frac{1}{u_x} = \frac{\sigma_0}{\sigma_i}, \tag{5.10}$$

is the inverse of the restriction parameter u_x with σ_0 being the standard deviation of the pooled metric scores of all lighting scenarios and σ_i representing the standard deviation of the range restricted metric scores of the i^{th} lighting scenario. Thus, by applying the above equations, the range restriction in the independent variable between the different studies can easily be corrected and the overall comparison of the individual correlations becomes feasible.

Another attenuation factor is induced by idiosyncrasy in the preference ratings of the individual observers. According to Smet *et al.* [54], the reliability $r_{yy,i}$ in the observer ratings and, therefore, the degree of the resulting attenuation $\sqrt{r_{yy,i}}$ of the true correlation of the i^{th} lighting scenario can be estimated from the respective inter-observer variability measured in terms of STRESS by assuming that

$$r_{yy,i} = 1 - \frac{\mathrm{STRESS}_{\mathrm{inter},i}}{100} \tag{5.11}$$

can be treated as a systematic measurement error so that the correction for inter-observer idiosyncrasy giving a correlation value $r_i^{\mathrm{corr.}}$ closer to the true correlation is simply obtained by

$$r_i^{\mathrm{corr.}} = \frac{r_i}{\sqrt{r_{yy,i}}}. \tag{5.12}$$

Unfortunately, for most of the psychophysical studies summarized in the previous section, the individual observer ratings and, consequently, the corresponding inter-observer variability measures were not accessible. For this reason, a $\text{STRESS}_{\text{inter}}$ value of 35 is assumed here, which is the typical value of inter-observer variability obtained in various color discrimination studies [219, 220] and, therefore, considered to be a good approximation to be applied in the meta-correlation analysis.

Finally, the sample correlation r_i of the i^{th} lighting scenario should be corrected for statistical bias. With the sample correlation being a statistical estimator of the true population correlation, the impact of the observed bias is systematic and can consequently be captured to a good approximation by an attenuation multiplier. As proposed by Hunter and Schmdit [298], the best attenuation multiplier in a meta-correlation analysis for absolute (estimated) population correlations smaller than 0.7 is a linear attenuation factor given by

$$a_i = 1 - \frac{1}{2N_i - 1}, \tag{5.13}$$

which, as can be noticed, is independent of the population correlation. For absolute values larger than 0.7 the more accurate, nonlinear attenuation factor

$$a_i = 1 - \frac{1 - r_i^2}{2N_i - 1} \tag{5.14}$$

should be used. Please note that in the above equation the actual population correlation was replaced by its best estimator, i.e., the sample correlation r_i. The final correction for statistical bias is simply given by

$$r_i^{\text{bias}} = \frac{r_i}{a_i} \tag{5.15}$$

and should be the last artifact to be corrected for in a meta-correlation analysis (at least for correlations >0.7) since the attenuation multiplier a_i is directly related to the population correlation estimator r_i.

Applying all attenuation correction steps in consecutive order finally leads to the best estimate $\hat{\bar{r}}_{\text{c}}$ of the true correlation between the metric's predictions and the observers' preference ratings. Hence, the complete correction formula is given by

$$\begin{aligned} \hat{\bar{r}}_{\text{c}} &= \left(\sum_{i=1}^{k} N_i^{\text{opt.}} \frac{U_x r_i}{a_i^* \cdot \sqrt{r_{yy,i} \left((U_x^2 - 1)\, r_i^2 + 1 \right)}} \right) \left(\sum_{i=1}^{k} N_i^{\text{opt.}} \right)^{-1} \\ &= \left(\sum_{i=1}^{k} N_i^{\text{opt.}} \frac{r_i^*}{a_i^*} \right) \left(\sum_{i=1}^{k} N_i^{\text{opt.}} \right)^{-1}, \end{aligned} \tag{5.16}$$

where

$$a_i^* = \begin{cases} 1 - (2N_i - 1)^{-1} & \text{if } |r_i^*| < 0.7 \\ 1 - \left(1 - (r_i^*)^2 \right) (2N_i - 1)^{-1} & \text{otherwise} \end{cases} \tag{5.17}$$

is the modified version of Eqs. (5.13) and (5.14) with the statistical bias being the last attenuation factor to be corrected for. Furthermore, the heterogeneity estimator for the calculation of

$N_i^{\text{opt.}}$ must also be adjusted accordingly. The formula for the Q value consequently changes to

$$Q = \sum_{i=1}^{k} N_i \left(\frac{r_i^*}{a_i^*} - \bar{r}^* \right)^2 , \tag{5.18}$$

with

$$\bar{r}^* = \left(\sum_{i=1}^{k} N_i \frac{r_i^*}{a_i^*} \right) \left(\sum_{i=1}^{k} N_i \right)^{-1} \tag{5.19}$$

being the modification of Eq. (5.4) corrected for range restriction, idiosyncrasy, and statistical bias artifacts, respectively. For more details on the mathematical concept adopted here, the interested reader is referred to the textbook of Hunter and Schmidt [298], which provides further insights and discussions on all aspects of meta-correlation analysis as a powerful tool in various fields of statistics.

5.2.3 *Results of the Meta-correlation Analysis and Cross-comparison of Metric Predictions*

After having discussed the mathematical and conceptional design of the meta-correlation analysis to be applied in this thesis, Table. 5.4 summarizes the finally obtained weighted average artifact-corrected Spearman correlation coefficients providing a measure for the predictive performance of the various tabulated color quality metrics that should be compared with each other. The results for both kinds of lighting scenarios (metameric vs. multi-CCT) are presented separately. For the sake of completeness, the intermediate results of the single artifact correction steps are also shown for each case indicated by a convenient labeling given in the table's first column. As can be seen, the individual artifact corrections generally had a de-attenuating effect. In some cases though, especially when correcting for study heterogeneity only, a slight decrease in the correlation coefficients could be observed. However, the application of the complete correction formula given by Eq. (5.16), whose results are tabulated in the rows labeled by "All", eventually lead to a significant improvement in the correlations for all color quality metrics.

Table 5.4 – Comparison of the various predictive metric performances in terms of weighted average artifact-corrected Spearman correlation coefficients obtained from meta-analysis. The correlation coefficients were calculated for both metameric and multi-CCT lighting scenarios. The intermediate results of the subsequently applied artifact correction steps are also tabulated (top to bottom), where the term "bare-bones" indicates the correction for sampling error only, while the numbers 2, 3, and 4 represent the application of the correction formulae for study heterogeneity, range restriction, and observer idiosyncrasy, respectively.

	Metameric lighting scenarios																	
Corrections	R_a	R_f	GAI	GAI/R_a	Q_a	Q_f	Q_p	Q_g	R_{2012}	FCI	R_g	R_p	R_{CPI}	$R_{flatt.}$	R_m	$MCPI_{Gl.}$	$MCPI_{Ger.}$	$MCPI_{Ch.}$
Bare-bones	0.02	0.10	0.63	0.40	0.25	0.01	0.82	0.71	0.28	0.67	0.73	−0.34	0.78	0.49	0.76	0.76	0.68	0.69
# 2	0.06	0.17	0.60	0.43	0.24	0.05	0.77	0.69	0.34	0.67	0.71	−0.32	0.76	0.51	0.75	0.75	0.68	0.73
# 2, 3	0.03	0.18	0.66	0.51	0.29	0.01	0.81	0.67	0.36	0.66	0.71	−0.36	0.77	0.59	0.79	0.91	0.81	0.79
# 2, 3, 4	0.05	0.23	0.72	0.61	0.35	0.03	0.81	0.74	0.44	0.72	0.79	−0.42	0.86	0.69	0.88	0.98	0.86	0.81
All	0.05	0.24	0.72	0.61	0.35	0.03	0.83	0.74	0.44	0.72	0.79	−0.42	0.86	0.69	0.88	0.98	0.86	0.83
	Multi-CCT lighting scenarios																	
Corrections	R_a	R_f	GAI	GAI/R_a	Q_a	Q_f	Q_p	Q_g	R_{2012}	FCI	R_g	R_p	R_{CPI}	$R_{flatt.}$	R_m	$MCPI_{Gl.}$	$MCPI_{Ger.}$	$MCPI_{Ch.}$
Bare-bones	−0.28	−0.28	0.40	0.70	−0.27	−0.27	−0.06	−0.35	0.18	−0.35	−0.34	0.54	0.68	0.68	−0.16	0.75	0.64	0.78
# 2	−0.27	−0.27	0.41	0.67	−0.26	−0.26	−0.05	−0.33	0.18	−0.34	−0.32	0.53	0.67	0.67	−0.14	0.73	0.62	0.75
# 2, 3	−0.48	−0.44	0.43	0.78	−0.42	−0.43	−0.07	−0.48	0.34	−0.43	−0.47	0.64	0.70	0.66	−0.21	0.93	0.90	0.90
# 2, 3, 4	−0.52	−0.50	0.50	0.82	−0.48	−0.49	−0.08	−0.53	0.36	−0.48	−0.51	0.72	0.77	0.74	−0.22	0.95	0.92	0.92
All	−0.53	−0.51	0.51	0.82	−0.49	−0.50	−0.08	−0.54	0.36	−0.49	−0.52	0.72	0.78	0.76	−0.23	0.95	0.92	0.92

In order to gain a better overview, the weighted average artifact-corrected Spearman correlation coefficients for each of the two different kinds of lighting scenarios were plotted in Fig. 5.1 together with their respective 95 % confidence intervals (CIs). By comparing the extent and the range of the various CIs, it should be obvious that the different color quality metrics vary considerably in their predictive power. To determine which of them is best suited for evaluating color rendering properties of white light sources with respect to visual appreciation, the observed performance differences are examined for significance in a series of cross-comparisons following the procedure described by Zou [302], who discussed the principles of using CIs to directly compare such kind of correlations.

Let us for example consider two different artifact-corrected Spearman correlation coefficients $\hat{\bar{r}}_{c,i}$ and $\hat{\bar{r}}_{c,j}$ supposed to give the best estimate of the true correlation with the observers' preference ratings for two different color quality metrics i and j which were arbitrarily chosen from Table 5.4. By applying a Fisher's z-transformation [303, 304], the corresponding 95 % CIs (l_i, u_i) and (l_j, u_j) for each correlation coefficient are given by

$$\begin{aligned} l_i &= \frac{\exp(2l_i^*) - 1}{\exp(2l_i^*) + 1}, \\ u_i &= \frac{\exp(2u_i^*) - 1}{\exp(2u_i^*) + 1}, \end{aligned} \tag{5.20}$$

and

$$\begin{aligned} l_j &= \frac{\exp(2l_j^*) - 1}{\exp(2l_j^*) + 1}, \\ u_j &= \frac{\exp(2u_j^*) - 1}{\exp(2u_j^*) + 1}, \end{aligned} \tag{5.21}$$

where

$$\begin{aligned} l_i^* &= \frac{1}{2} \cdot \ln\left(\frac{1 + \hat{\bar{r}}_{c,i}}{1 - \hat{\bar{r}}_{c,i}}\right) - 1.96 \cdot \sqrt{\frac{1}{\overline{N} - 3}}, \\ u_i^* &= \frac{1}{2} \cdot \ln\left(\frac{1 + \hat{\bar{r}}_{c,i}}{1 - \hat{\bar{r}}_{c,i}}\right) + 1.96 \cdot \sqrt{\frac{1}{\overline{N} - 3}}, \end{aligned} \tag{5.22}$$

and

$$\begin{aligned} l_j^* &= \frac{1}{2} \cdot \ln\left(\frac{1 + \hat{\bar{r}}_{c,j}}{1 - \hat{\bar{r}}_{c,j}}\right) - 1.96 \cdot \sqrt{\frac{1}{\overline{N} - 3}}, \\ u_j^* &= \frac{1}{2} \cdot \ln\left(\frac{1 + \hat{\bar{r}}_{c,j}}{1 - \hat{\bar{r}}_{c,j}}\right) + 1.96 \cdot \sqrt{\frac{1}{\overline{N} - 3}}, \end{aligned} \tag{5.23}$$

with $\overline{N}$ denoting the average observer number of all studies used to calculate the respective correlation coefficients. Assuming independent correlations, it was shown by Zou [302] that the 95 % CI (L, U) for the difference $\hat{\bar{r}}_{c,i} - \hat{\bar{r}}_{c,j}$ between both correlations can be derived as

$$\begin{aligned} L &= \hat{\bar{r}}_{c,i} - \hat{\bar{r}}_{c,j} - \sqrt{\left(\hat{\bar{r}}_{c,i} - l_i\right)^2 + \left(u_j - \hat{\bar{r}}_{c,j}\right)^2}, \\ U &= \hat{\bar{r}}_{c,i} - \hat{\bar{r}}_{c,j} + \sqrt{\left(u_i - \hat{\bar{r}}_{c,i}\right)^2 + \left(\hat{\bar{r}}_{c,j} - l_j\right)^2}, \end{aligned} \tag{5.24}$$

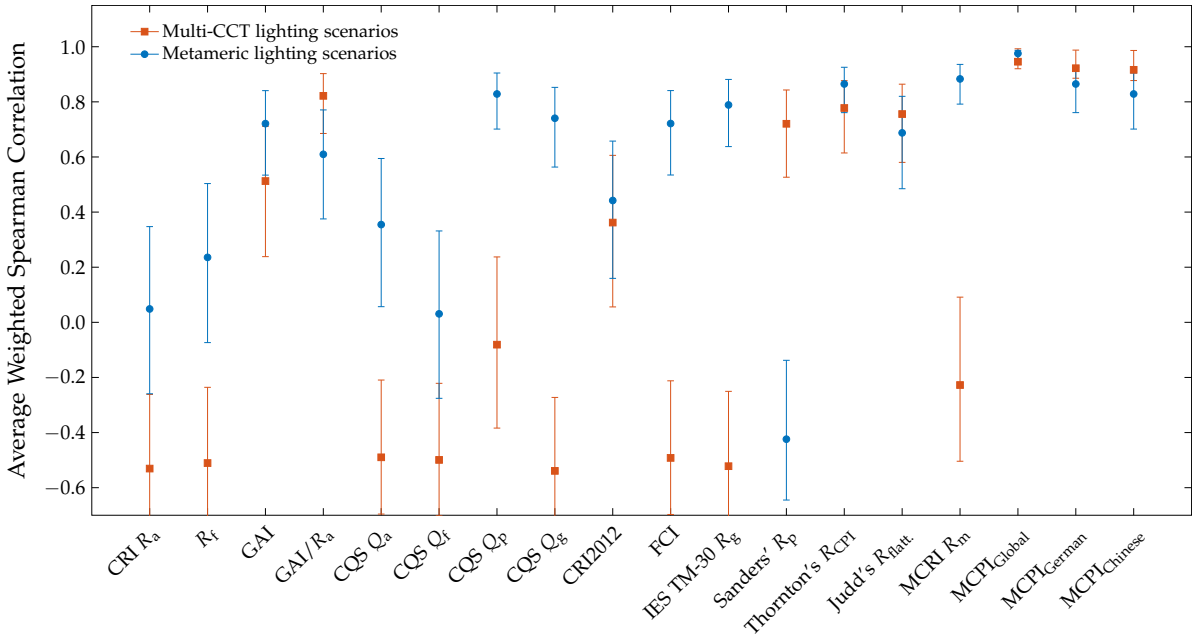


Figure 5.1 – Comparison of the predictive performance of various color quality metrics expressed by artifact-corrected Spearman correlation coefficients obtained from meta-correlation analysis which describe the individual metric's ability of correctly ranking light sources in terms of visual appreciation based on human observers' preference ratings. The indicated errorbars represent corresponding 95 % confidence intervals. Metameric as well as multi-CCT lighting scenarios have been analyzed separately.

which provides a direct method for hypothesis testing when comparing metric correlations: If the CI of the correlation difference does not contain zero, the null hypothesis of equal correlations, i.e., $H_0 : \hat{\bar{r}}_{c,i} - \hat{\bar{r}}_{c,j} = 0$, must be rejected at a 5 % significance level. In Table. 5.5, the CI bounds of the performed CI tests on the correlation differences of the cross-compared color quality metrics are summarized for both kinds of lighting scenarios. Please note that only the CI bounds closest to zero are shown. Bold numbers indicate statistically significant differences.

METAMERIC LIGHTING SCENARIOS: Considering first only the metameric lighting scenarios, it can be seen that the newly proposed $MCPI_{Global}$ metric offers the best predictive performance among all color quality metrics. As mentioned before, the weighting factors of Eq. (5.3) were determined directly from the meta-correlation analysis to account for the varying degree of importance of the familiar test objects' colors in the assessment of the visual appreciation and the general color preference for an arbitrary test illuminant. For this purpose, a global search algorithm for nonlinear objective functions described by Ugray *et al.* [305] was applied here in order to find a set of weighting factors that maximizes the corresponding artifact corrected Spearman correlation coefficient. The results of this optimization are summarized in Table 5.6.

With these optimized weighting factors, the overall artifact corrected correlation with the observers' color preference ratings was $\hat{\bar{r}}_{c,Global} = 0.976^{+0.011}_{-0.021}$, which is the largest correlation value among all compared color quality metrics. Furthermore, having a look at the lower grey-shaded triangle of Table. 5.5, which summarizes the results of the CI hypothesis tests for the metameric lighting scenarios, reveals that the $MCPI_{Global}$ metric performs significantly better in predicting visual appreciation than any of the other color quality metrics including Smet *et al.*'s MCRI which is also based on the evaluation of memory colors but, as discussed

in the previous chapters, makes use of an arguable test object selection and a less sophisticated color space neglecting the impact of different adaptation conditions on observers' color preference ratings.

However, it should be emphasized that even though significance could be confirmed, the tabulated CI bound of -0.04 for the correlation difference between the MCRI with $\hat{\hat{r}}_{\mathrm{c,MCRI}} = 0.883^{+0.053}_{-0.091}$ and the $\mathrm{MCPI}_{\mathrm{Global}}$ is quite small, which indicates that the difference in performance between these two metrics is only little pronounced for metameric lighting scenarios. Other metrics with slightly larger but still rather small CI bounds are Thornton's CPI with $\hat{\hat{r}}_{\mathrm{c,CPI}} = 0.865^{+0.061}_{-0.104}$, the Q_{p} metric with $\hat{\hat{r}}_{\mathrm{c},Q_{\mathrm{p}}} = 0.829^{+0.076}_{-0.127}$ as well as both cultural specific MCPIs which were calculated using the same weighting factors as optimized for $\mathrm{MCPI}_{\mathrm{Global}}$ giving $\hat{\hat{r}}_{\mathrm{c,Chinese}} = 0.830^{+0.076}_{-0.126}$ and $\hat{\hat{r}}_{\mathrm{c,German}} = 0.865^{+0.061}_{-0.104}$ for the Chinese and German average observer, respectively. Hence, with each of the cultural specific MCPIs showing comparably high correlations with barely significant differences to the global reference, it can be concluded that the impact of the cultural background (at least for the two different observer groups discussed in this thesis) on the metrics' predictive performance is most likely of no practical importance. This is basically in accordance to the findings of Sec. 4.4.3 where the effect size of the inter-cultural variations in the color appearance ratings of Chinese and German observers was reported to be quite small in general and much smaller than the impact of the adapted white point. Moreover, no significant differences between the memory color centers of Chinese and German observers could be found.

Therefore, indication is given that it would be sufficient to propose a single, universally valid MCPI that is capable of well predicting the rank order of metameric light sources with respect to their visual appreciation representing an average global observer. With the predictive performance of both cultural-specific MCPIs being comparable to the (slightly better) performance of the global MCPI, the latter can be considered as a good approximation to a globally valid color quality metric inducing only minor errors in the absolute level of the predicted results compared to those obtained for the cultural-specific MCPIs, which however are considered to be negligible in practice. Besides being applied to evaluate the color rendering properties of existing light sources, such a universally valid preference metric can also be used for the spectral optimization of new light sources which are intended to achieve a high user acceptability by providing excellent lighting quality for a globally oriented market. Even though cultural-specific fine-tuning might be feasible for certain regions when adopting the dedicated MCPI definitions, the small effect sizes of the inter-cultural variations reported for the color appearance ratings of familiar objects as well as the observed barely significant differences in the predictive performance of these metrics suggest that such a fine-tuning would most likely be of no practical importance for lighting designers so that an optimization using the globally valid approximation should be sufficient for most applications.

These findings are further emphasized when attempting to derive a new set of weighting factors intended to maximize the correlations with the visual ratings for each of the cultural-specific MCPIs separately. In both cases, the overall artifact-corrected correlations between the metrics' predictions and the observers' preference ratings of metameric lighting scenarios can be increased by applying the optimized cultural-specific weighting factors given in Table 5.6. The new correlation values read $0.976^{+0.012}_{-0.020}$ and $0.979^{+0.010}_{-0.018}$ for the Chinese and German average observer, respectively. Even though the CI test method reveals a statistically significant improvement over the results obtained for the original weighting factors at a 5 % significance level, the resulting lower CI bounds are still quite small so that the argumentation of the last paragraph also holds here leading to the conclusion that these slight improvements basically would be of no practical value. Furthermore, no significant differences in the

Table 5.5 – Overview of the results of the cross-comparison confidence interval test of Zou [302] intended to examine the predictive performance of the various color quality metrics for significant differences adopting a 5 % significance level. The null hypothesis $H_0: \hat{\bar{r}}_{c,i} - \hat{\bar{r}}_{c,j} = 0$ tests whether or not the observed correlations of two distinct metrics i and j are equal. The tabulated values are the confidence interval bounds of the correlation differences closest to zero, where bold numbers indicate significant differences in case that zero was not included within the confidence interval bounds. Results for the metrics' correlation with observers' preference ratings in metameric and multi-CCT lighting scenarios are shown in the lower and upper triangle of the table, respectively.

	R_a	R_f	GAI	GAI/R_a	Q_a	Q_f	Q_p	Q_g	R_{2012}	FCI	R_g	R_p	R_{CPI}	$R_{flatt.}$	R_m	$MCPI_{Gl.}$	$MCPI_{Ger.}$	$MCPI_{Ch.}$
R_a		0.31	**−0.66**	**−1.05**	0.30	0.31	**−0.04**	−0.32	**−0.49**	0.30	0.32	**−0.92**	**−0.99**	**−0.97**	0.08	**−1.20**	**−1.18**	**−1.17**
R_f	0.24		**−0.64**	**−1.03**	0.32	0.33	**−0.02**	−0.30	**−0.46**	0.32	−0.33	**−0.89**	**−0.97**	**−0.94**	0.11	**−1.18**	**−1.15**	**−1.14**
GAI	**−0.32**	**−0.16**		**−0.07**	**0.61**	**0.62**	**0.17**	**0.67**	−0.22	**0.61**	**0.65**	0.07	**−0.01**	0.02	**0.32**	**−0.23**	**−0.20**	**−0.19**
GAI/R_a	**−0.18**	**−0.02**	−0.13		**1.00**	**1.01**	**0.56**	**1.06**	**0.18**	**1.00**	**1.04**	−0.08	−0.12	−0.11	**0.70**	**−0.03**	0.00	0.01
Q_a	0.12	0.28	**0.06**	−0.08		−0.34	0.00	−0.29	**−0.44**	−0.35	−0.31	**−0.87**	**−0.94**	**−0.91**	0.13	**−1.15**	**−1.12**	**−1.12**
Q_f	−0.41	−0.23	**0.34**	**0.20**	−0.10		**−0.01**	−0.30	**−0.45**	0.34	−0.32	**−0.88**	**−0.95**	**−0.93**	0.12	**−1.16**	**−1.14**	**−1.13**
Q_p	**−0.46**	**−0.30**	0.07	**−0.01**	**−0.20**	**−0.47**		**0.05**	**−0.00**	−0.00	**0.03**	**−0.43**	**−0.50**	**−0.47**	−0.29	**−0.70**	**−0.68**	**−0.67**
Q_g	**−0.34**	**−0.18**	0.20	0.11	**−0.09**	**−0.36**	−0.08		**−0.50**	0.29	0.31	**−0.93**	**−1.00**	**−0.98**	0.07	**−1.21**	**−1.19**	**−1.18**
R_{2012}	0.02	0.18	−0.00	−0.15	0.28	0.00	**0.14**	**0.02**		**0.44**	**0.47**	**−0.05**	**−0.12**	**−0.09**	**0.15**	**−0.34**	**−0.31**	**−0.30**
FCI	**−0.32**	**−0.16**	−0.22	0.14	**−0.06**	**−0.34**	−0.07	−0.19	0.01		−0.31	**−0.87**	**−0.95**	**−0.92**	0.13	**−1.15**	**−1.13**	**−1.12**
R_g	**−0.41**	**−0.25**	0.13	0.04	**−0.15**	**−0.42**	−0.12	0.14	**−0.08**	0.13		**−0.91**	**−0.98**	**−0.95**	0.09	**−1.19**	**−1.16**	**−1.16**
R_p	**0.04**	**0.23**	**0.80**	**0.66**	**0.36**	**0.03**	**0.93**	**0.82**	**0.46**	**0.80**	**0.88**		0.15	0.18	**0.57**	**−0.09**	**−0.06**	**−0.05**
R_{CPI}	**−0.50**	**−0.34**	0.02	**−0.06**	**−0.25**	**−0.52**	0.09	0.03	**−0.18**	0.02	0.06	**−0.98**		−0.17	**0.65**	**−0.06**	**−0.03**	**−0.02**
$R_{flatt.}$	**−0.28**	**−0.12**	−0.19	0.18	**−0.02**	**−0.29**	−0.04	−0.17	0.05	−0.20	−0.10	**−0.75**	**0.01**		**0.62**	**−0.07**	**−0.04**	**−0.03**
R_m	**−0.52**	**−0.36**	**−0.01**	**−0.09**	**−0.27**	**−0.54**	0.06	0.00	**−0.21**	**−0.01**	0.04	**−1.00**	0.09	**−0.03**		**−0.85**	**−0.82**	**−0.82**
$MCPI_{Gl.}$	**−0.63**	**−0.47**	**−0.13**	**−0.20**	**−0.38**	**−0.64**	**−0.07**	**−0.12**	**−0.32**	**−0.13**	**−0.09**	**−1.10**	**−0.05**	**−0.15**	**−0.04**		−0.04	−0.03
$MCPI_{Ger.}$	**−0.50**	**−0.34**	0.02	**−0.06**	**−0.25**	**−0.52**	0.09	0.03	**−0.18**	0.01	0.06	**−0.98**	0.12	**−0.01**	−0.09	**0.05**		−0.07
$MCPI_{Ch.}$	**−0.46**	**−0.30**	0.07	**−0.01**	**−0.20**	**−0.47**	0.15	0.08	**−0.14**	0.07	0.12	**−0.93**	−0.09	0.04	−0.06	**0.07**	−0.09	

predictive power of the optimized cultural-specific MCPIs and the global reference could be observed making them interchangeable for the cross-regional evaluation of the visual appreciation of metameric lighting scenarios, which additionally confirms the postulate of using the $MCPI_{Global}$ as an approximation to a single, universally valid memory-based color quality metric.

As mentioned above, both Thornton's CPI as well as Smet *et al.*'s MCRI exhibit a good-to-excellent correlation with observers' preference ratings for metameric lighting scenarios approaching the correlation observed for the best performing metric ($MCPI_{Global}$), which indicates good but still significantly weaker predictive power for both metrics. In contrast, Sanders' R_p and Judd's $R_{flatt.}$, which are the remaining two memory- or preference-related color quality metrics, only show poor-to-moderate ($\hat{\bar{r}}_{c,Sanders} = -0.417^{+0.288}_{-0.223}$) and moderate-to-good correlations ($\hat{\bar{r}}_{c,Judd} = 0.688^{+0.133}_{-0.203}$), respectively. Whereas the latter at least succeeds to predict a positive trend in the rank order so that a more preferred light source also tends to result in a larger metric value, which from the author's point of view should be a key feature of any properly devised color quality metric, the former must be considered to fail completely for metameric lighting scenarios. This is not just because of its generally poor predictive performance in terms of visual appreciation but also because of its non-intuitive, opposing light source ranking.

Possible explanations for the poor performance of Sanders' R_p for these kind of lighting scenarios are manifold: First of all, in the definition of his calculation scheme, Sanders made use of the perceptually non-uniform CIE 1931 (x, y) chromaticity space adopting the CIE 2° standard observer which is known to result in relatively large errors in the blue part of the spectrum [306, 307] expected to cause considerable deviations from visual color matches for many of the LED light sources used in the visual experiments reported in the compilation of metameric studies of Sec. 5.2.1. Second, a critical mismatch between the sample set defined by Sanders (only reddish to yellowish samples) and the experimental objects sets used in the various studies on color preference can be identified. Hence, with Sanders' R_p metric lacking

Table 5.6 – Summary of the optimized weighting coefficients introduced in Eq. (5.3) for the definition of the global and both cultural-specific MCPI color quality metrics. The individual p_i values represent from left to right the weighting coefficients for the test objects of Asian skin, banana, blueberry, blue jeans, broccoli, butternut, carrot, Caucasian skin, concrete, green salad, red cabbage, and red rose. For each kind of lighting scenario (metameric vs. multi-CCT) the appropriate set of weighting coefficients has been determined by running a global search optimization algorithm on the meta-correlation analysis in order to maximize the resulting correlations with the respective observers' preference ratings.

	Metameric lighting scenarios											
Type	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}
Global	0.001 601	0.001 689	0.001 624	0.001 658	0.736 855	0.001 655	0.001 692	0.001 547	0.229 577	0.000 219	0.001 038	0.020 844
German	0.000 123	0.002 870	0.000 899	0.001 725	0.915 560	0.003 100	0.014 343	0.000 109	0.002 164	0.000 954	0.000 362	0.057 790
Chinese	0.002 971	0.007 384	0.006 925	0.006 015	0.726 700	0.006 835	0.005 173	0.002 482	0.215 285	0.002 584	0.003 707	0.013 939
	Multi-CCT lighting scenarios											
Type	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}
Global	0.246 416	0.003 331	0.023 070	0.025 772	0.231 410	0.078 222	0.179 669	0.041 295	0.019 311	0.091 012	0.048 690	0.011 802
German	0.200 039	0.014 890	0.065 191	0.011 156	0.128 922	0.009 938	0.313 526	0.021 616	0.073 243	0.089 176	0.072 061	0.000 243
Chinese	0.129 331	0.043 434	0.023 668	0.037 430	0.220 705	0.078 088	0.067 109	0.094 445	0.028 116	0.162 172	0.090 688	0.024 814

of full hue coverage, it is not surprising that such a bias would yield relatively poor results, especially for metameric lighting scenarios where the white point of adaptation is kept fix and changes in the color appearance of the presented test objects, which in most cases are not well represented by Sanders' limited set of samples, are the only criteria for rating a certain light source lower or higher in preference. A third reason for the poor predictive performance of Sanders' R_p metric might be the use of an outdated chromatic adaptation transform of the translational Judd-type (see Sec. 3.2.1). Although an update to the more sophisticated CAT02 adaptation transform (with $D = 1$) as described in Sec. 2.2.2 leads to an outstanding improvement in the artifact-corrected correlation result giving $\hat{\hat{r}}^*_{\mathrm{c,Sanders}} = 0.460^{+0.261}_{-0.204}$, Sanders' R_p still exhibits a rather poor-to-moderate predictive performance in terms of visual appreciation for metameric lighting scenarios. This indicates that the remaining two drawbacks discussed above are likely to be the more critical problems in the metric's definition.

As mentioned previously, Judd's flattery index $R_{\mathrm{flatt.}}$ and Thornton's CPI both exhibit a much better correlation with the observers' color preference ratings for metameric visual experiments than Sanders' R_p metric, even though they all made use of the same outdated translational chromatic adaptation transform. However, in contrast to the concept of Sanders' R_p, a more perceptually uniform color space as well as a set of test samples well distributed around the hue circle were adopted for the construction of the other two preference-based color quality metrics. This basically confirms the findings of the preceding paragraph emphasizing the importance of full hue coverage and the use of an adequately chosen perceptually uniform color space to achieve a good-to-excellent metric performance. Please note that in the development of the newly proposed MCPI special care was taken to fulfill these requirements.

When comparing Judd's flattery index $R_{\mathrm{flatt.}}$ and Thornton's CPI, it can be further seen from Table 5.5 and Fig. 5.1 that the latter performs somewhat better in terms of visual appreciation for metameric lighting scenarios than the former. However, with the reported CI bound of 0.01 being very close to zero, indication is given that the observed differences between the predictive performance of both metrics are barely significant. As discussed in Secs. 3.2.2 and 3.2.3, the major differences between both color quality metrics are i) the magnitude of the preference shifts added to the chromaticities of the Munsell test samples as perceived under reference illumination in order to obtain the preferred chromaticities and ii) the weighting factors for the individual test samples only introduced by Judd. The rest

of both concepts is more or less the same. Hence, by setting all weighting factors to unity and, at the same time, using the full magnitude of the preferred chromaticity shifts for the computation of $R_{\text{flatt.}}$, the eventually resulting correlation increases to $\hat{\hat{r}}^{*}_{\text{c,Judd}} = 0.879^{+0.059}_{-0.099}$ which is even slightly higher than the correlation reported for Thornton's CPI and, therefore, also comparable to the predictive power of the best performing memory-based color quality metrics – at least for metameric lighting scenarios. A conformable increase in correlation obtained by adjusting Judd's $R_{\text{flatt.}}$ accordingly was also observed in a similar meta-correlation analysis performed by Smet *et al.* [54] who concluded that the slightly higher correlation value obtained for Judd's $R_{\text{flatt.}}$ when compared to Thornton's CPI was most likely because of the two extra color samples providing additional coverage of the hue circle.

Furthermore, as in the case of Sanders' R_{p}, the replacement of the outdated translational chromatic adaptation transform by the CAT02 may additionally increase the correlations of both metrics. Doing so, one obtains correlations of $\hat{\hat{r}}'_{\text{c,Judd}} = 0.907^{+0.049}_{-0.087}$ and $\hat{\hat{r}}'_{\text{c,CPI}} = 0.882^{+0.054}_{-0.093}$ for Judd's $R_{\text{flatt.}}$ and Thornton's CPI, respectively. As can be seen, the positive effect is not as large as for Sanders' R_{p}, which according to Smet *et al.* [54] can be explained by considering the magnitude of the adaptive shifts required in the metric calculations. Whereas for the computation of Sanders' R_{p} either CIE illluminant B or C is used requiring in general large adaptive shifts when evaluating arbitrary test light sources with chromaticites that considerably differ from these two possible references, the $R_{\text{flatt.}}$ and CPI metrics, on the other hand, make use of reference illuminants that exhibit the same CCT as the test light source so that in these cases the adaptive shifts typically required and, therefore, the error introduced by the outdated translational chromatic adaptation transform will be much smaller than in the former case.

Coming back to Fig. 5.1, it is clear that the newly proposed $\text{MCPI}_{\text{Global}}$ outperforms not just the other memory- and preference-based color quality metrics but also all of the remaining measures. With respect to the predictive power of visual appreciation for metameric lighting scenarios, the best performing color quality metrics that are not memory- or preference-related are those that can be summarized as either gamut- (Q_{g}, R_{g}, GAI, FCI, mixed GAI/R_{a}) or chroma-enhancement-based (Q_{p}) metrics. Both categories show correlation values comparable to those observed for the preference-based metrics and Smet *et al.*'s MCRI. This is confirmed by the CI test where, as summarized in Table 5.5, cross-comparisons allow for drawing the conclusion of no or only barely significant differences between the predictive performances of all these metrics. Just the global MCPI performs significantly much better.

The overall lowest group performance for metameric lighting scenarios, on the other hand, can be noticed for the fidelity (CRI R_{a}, R_{f}, Q_{f}, CRI2012) and the fidelity-oriented (Q_{a}) metrics, all of which perform significantly worse than the memory-and preference-based approaches (with the exception of Sanders' R_{p}) and in most cases also worse than the gamut- and chroma-enhancement-based alternatives. Obviously, color fidelity metrics are not intended to measure color rendering properties of light sources in terms of visual appreciation. Indeed, they were originally developed to evaluate how similar an arbitrary test light source is to a certain reference standard of the same CCT and, therefore, should only be used for this purpose. This is further emphasized when looking at Table 5.4 from which it can be noticed that both the CIE R_{a} and the CQS Q_{f} exhibit correlation values close to zero indicating that both metrics completely fail to model the observers' preference ratings properly. Consequently, such metrics should not be used for the evaluation and optimization of light sources and luminaires for achieving high user-acceptability and visual appreciation. Unfortunately, the application of, in particular, the CIE R_{a} for such purposes is still common practice in the

industry emphasizing the necessity of establishing a new and better standard in the near future.

As reported by Smet *et al.* [54] and which can also be observed here, the predictive performance for visual appreciation tends to increase the more weight a color quality metric puts on chroma enhancement or gamut expansion which both are commonly considered to increase preference up to a certain limit [19]. Regarding for example the role of chroma enhancement, a clear tendency can be derived from the four different CQS indices which show a considerable gain in performance as the reward for chroma enhancement in the respective calculation scheme increases. With Q_f explicitly excluding the saturation factor originally introduced in the Q_a calculation for not penalizing moderate increases in chroma caused by a test light source, it is not surprising that the former generally performs worse than the latter in modeling visual appreciation for metameric lighting scenarios. A further improvement is obtained when applying the Q_p and Q_g metrics which both additionally reward light sources for increasing object chroma. As expected, this leads to a significantly better correlation with the preference rating results than obtained for the other two more fidelity-oriented CQS indices.

Here, it should be noted that the Q_g index actually falls into the category of gamut-based color quality metrics rather than offering pure chroma enhancement. As mentioned earlier, comparably good predictive performance for metameric lighting scenarios can be found for the related gamut-based alternatives R_g, GAI, and FCI. However, in contrast to Q_p or the preference- and memory-based approaches, the gamut expansion metrics discussed here do not contain an upper limit for visually allowed chroma augmentation. It is commonly known and has been proven in various studies [308–312] that the oversaturation of object colors beyond a certain degree has a negative impact on the perceived color quality and visual appreciation. Hence, the fact that the above mentioned gamut expansion approaches lack the possibility of setting a limit to the potential chroma enhancement caused by a test light source might be an explanation for their on average slightly worse predictive performance observed from Fig. 5.1 compared to the performance of the non-fidelity metrics that do include such a limit, either by defining certain reference chromaticities with more or less fixed values of increased saturation (R_p, $R_{flatt.}$, CPI, MCRI R_m, MCPI) or by explicitly giving a limit for the maximally allowed chroma enhancement (CQS Q_p). A special case is the combined GAI/R_a metric which basically tries to counterbalance too large increases in chroma by penalizing induced deviations from the unsaturated reference chromaticities. As can be seen from Fig. 5.1, this special concept leads to a predictive performance for metameric lighting scenarios ranging between the performance of those color quality metrics that reward chroma enhancement and of those that are only fidelity-based or -oriented.

MULTI-CCT LIGHTING SCENARIOS: In contrast to the evaluation of the metameric visual data where, with the exception of the fidelity-based approaches and Sanders' R_p, most of the color quality metrics performed quite well in properly ranking the test light sources in terms of observers' visual appreciation indicated by relatively large positive Spearman correlation coefficients, the situation is completely different in case of considering multi-CCT lighting scenarios. Similar to Sanders R_p in the metameric case, most of the investigated color quality metrics confound the rank order of the light sources resulting in poor-to-moderate negative correlations. With the exception of the FCI, which uses a constant reference illuminant (CIE D65) for calculating the relative gamut area spanned by the apparent chromaticities of four test samples, and Smet *et al.*'s MCRI, which uses reference chromaticities based on memory colors, all of these poorly performing metrics adopt a reference illuminant of the same CCT

as the test light source (CRI R_a, R_f, all four CQS indices, R_g). However, in a multi-CCT scenario the values of such measures calculated for different CCTs are not really comparable because by definition they are correlated to different reference light sources [294, 295] and, therefore, completely fail to model the general trend of visual preference ratings of observers assessing a series of multi-CCT illuminants. The only exception thereto is the CRI2012 metric which still shows a very poor but at least non-negative correlation of $\hat{\bar{r}}_{c,CRI2012} = 0.362^{+0.244}_{-0.306}$. This slightly better performance than the rest of the color quality metrics whose calculation schemes are based on a CCT-dependent reference illuminant might be explained by the adoption of the CIECAM02-UCS including the application of the state-of-the-art CAT02 chromatic adaptation transform, the usage of the CIE 10° standard observer, and a sophisticated test sample selection providing a much better hue coverage in the definition of the CRI2012 metric.

Compared to the color quality metrics discussed in the last paragraph, significantly better results are in principle observed for the preference- and memory-based approaches as well as for both variations of the GAI metric which, similar to the FCI approach, make use of a fixed reference illuminant for the calculation of the relative gamut area. With correlations of $\hat{\bar{r}}_{c,GAI} = 0.513^{+0.199}_{-0.274}$ and $\hat{\bar{r}}_{c,GAI/R_a} = 0.822^{+0.081}_{-0.136}$ obtained for the GAI and the GAI/R_a metric, cross-comparison CI tests summarized in the upper triangle of Table 5.5 revealed no or only barley significant differences in their predictive power for multi-CCT lighting scenarios when compared to the preference-based color quality metrics of Sanders, Thornton, and Judd showing correlation coefficients of $\hat{\bar{r}}'_{c,Sanders} = 0.720^{+0.123}_{-0.194}$, $\hat{\bar{r}}'_{c,CPI} = 0.778^{+0.099}_{-0.163}$, and $\hat{\bar{r}}'_{c,Judd} = 0.755^{+0.108}_{-0.175}$, respectively. Please note that for the computation of the indicated correlation values the best performing version of each preference-based metric as reported for the metameric lighting scenarios has been used here, denoted by the primed labeling, i.e., i) in each case the outdated translational adaptation transform included in the original metric definitions has again been replaced by the more sophisticated CAT02 formula and ii) for Judd's $R_{flatt.}$ the weighting factors and the magnitudes of the preferred chromaticity shifts have been adjusted accordingly to match those of Thornton's CPI.

As can be seen from the reported Spearman correlation coefficients, all of the so-improved preference-based color quality metrics exhibit moderate-to-good predictive performance in terms of visual appreciation for multi-CCT lighting scenarios. Somewhat surprising is to obtain such a good result even for Sanders R_p which despite using the more adequate chromatic adaptation transform still performed rather poor in the metameric case. This significant performance improvement might be due to the fact that the assessment of the color appearance of certain test objects in a multi-CCT scenario always involves the adaptation to a changing illumination white point. Even though observers of such studies are usually asked to wait some time before giving their ratings in order to be fully adapted or to close their eyes during the change of the light source to be assessed, there is still some cognitive effect caused by the first glance a new lighting situation is perceived which, indisputably, has a non-negligible impact on the observers' preference ratings. Hence, with the impact of the perceived white point increasing, the lack in hue coverage of Sanders' test samples used for the R_p calculation is less severe for predicting visual appreciation in multi-CCT than in metameric lighting scenarios.

Very poor performance, on the other hand, is observed for Smet *et al.*'s MCRI showing a Spearman correlation coefficient of $\hat{\bar{r}}_{c,MCRI} = -0.227^{+0.319}_{-0.277}$ which is comparable to those of the worst performing color quality metrics for multi-CCT lighting scenarios. A possible explanation might be that in cases where chromatic re-adaptation between two subsequent ratings becomes necessary due to a change in the white point of the illumination, some

additional cognitive process could be triggered which might cause a considerable shift in the recalled chromaticities of the memory color centers when the familiar test objects used for constructing the memory-based metric would be assessed under such a consecutively changing illumination. This potentially induced error of chromatic adaptation becomes increasingly more severe the larger the CCT of the test light source in a multi-CCT scenario deviates from the reference illumination under which the memory color centers have originally been determined. For metameric lighting scenarios, on the other hand, these kind of errors also exists, however, they can be assumed to be approximately constant among the different test light sources which have more or less the same white points so that despite those errors of chromatic adaptation the MCRI metric is still capable of getting the rank order correctly leading to the observed good-to-excellent correlation for metameric light sources, even though absolute deviations in terms of visual appreciation might also be expected for these cases.

In order to get at least partly hold of this problem the newly proposed MCPI metric was devised in such a way (see Sec. 5.1) that, similar to Sanders' R_p, the impact of the adapted white point on the assessment of memory colors is approximated by the implementation of a decision algorithm which, dependent on the CCT of the test light source, chooses a more appropriate set of similarity distribution functions being assumed to give a better estimate for the test objects' true memory color for that specific adaptation condition. In addition to this conceptional improvement, a new set of weighting factors specifically designated for multi-CCT lighting scenarios was also optimized with the resulting p_i values being given in Table 5.6. The intention here is to counterbalance the introduced errors of chromatic adaptation discussed above, which are expected to be more severe for certain test colors than for others, by some adequately chosen color weighting factors giving less weight to the colors that exhibit larger errors.

As can be seen from Fig. 5.1, the eventually calculated MCPIs exhibit superior correlations with the observers' preference ratings giving correlation coefficients of $\hat{\hat{r}}_{\mathrm{c,Global}} = 0.946^{+0.047}_{-0.026}$, $\hat{\hat{r}}_{\mathrm{c,German}} = 0.922^{+0.066}_{-0.036}$, and $\hat{\hat{r}}_{\mathrm{c,Chinese}} = 0.916^{+0.070}_{-0.039}$ for the global and the two cultural-specific MCPIs, respectively. Furthermore, analyzing the results of the CI cross-comparisons of Table 5.5 reveals that the MCPI metric offers significantly better predictive performance with respect to observers' preference than any of the other color quality metrics under inspection. Only for the GAI/R_a which performs best among all these metrics a slightly larger overlap is observed leading to non-significant differences in performance when compared to both cultural-specific MCPIs. However, like in the case of metameric lighting scenarios, a notable improvement in the correlations of the latter could be achieved by re-applying the global search algorithm to find for each of these cultural-specific MCPIs a new set of optimized weighting factors also given in Table 5.6. Hence, the updated correlation coefficients read $\hat{\hat{r}}^{*}_{\mathrm{c,German}} = 0.945^{+0.047}_{-0.027}$ and $\hat{\hat{r}}^{*}_{\mathrm{c,Chinese}} = 0.937^{+0.062}_{-0.031}$, respectively, which ultimately is significantly better than the performance of the GAI/R_a metric . Again, no significant differences in the predictive performance between the global and the cultural-specific MCPIs can be found neither before nor after the weighting factor optimization. Consequently, it can be concluded that even for multi-CCT lighting scenarios the $\mathrm{MCPI}_{\mathrm{Global}}$ metric seems to be a pretty good approximation to be used as a single, universally valid memory-based color quality metric.

5.3 SUMMARY (II)

In this chapter, an updated and improved memory-based color quality metric for the evaluation of the color rendering properties of white light sources in terms of visual appreciation denoted as memory color preference index MCPI has been proposed. It is based on the evaluation of the degree of similarity between the color appearance of certain familiar test objects rendered by an arbitrary test light source and their respective memory color representation. The degree of similarity is assessed by using a set of similarity distribution functions fitted to the results of previously conducted color appearance rating experiments of Chinese and German observers. The key features of the newly proposed color quality metric are the adoption of the perceptionally uniform CIECAM02-UCS as the working color space, the implementation of a CCT-based decision algorithm to choose a suitable set of similarity functions better approximating the impact of the adapted white point on the memory color assessments, and the introduction of some additional weighting factors allowing i) to model the varying importance of certain test colors in the evaluation of light sources with respect to color preference and ii) to counterbalance the metric errors in the memory color assessments introduced by chromatic adaptation.

Subsequently, a meta-correlation analysis considering the experimental data of a comprehensive selection of different visual studies investigating color preference ratings of human observers in both metameric and multi-CCT lighting scenarios was performed on a representative variety of different color quality metrics. It could be shown that the newly proposed MCPI metric significantly outperformed all alternative approaches showing a much better performance in terms of predicting visual appreciation of light sources. Furthermore, no or only barely significant differences between the global and the cultural-specific MCPI definitions could be found with respect to their predictive performance indicating that the former is a sufficiently good approximation to be used as a single, universally valid memory-based color quality metric which performs significantly better than the alternatives considered in this thesis.

REFERENCES OF CHAPTER V

[19] W. Davis and Y. Ohno, "Color quality scale," *Optical Engineering*, vol. 49, no. 3, p. 033 602, 2010.

[20] M. S. Islam, R. Dangol, M. Hyvärinen, *et al.*, "Investigation of User Preferences for LED Lighting in Terms of Light Spectrum," *Lighting Research & Technology*, vol. 45, no. 6, pp. 641–665, 2013.

[23] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Memory Colours and Colour Quality Evaluation of Conventional and Solid-State Lamps," *Optics Express*, vol. 18, no. 25, pp. 26 229–26 244, 2010.

[32] E. Hering, *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer Verlag, 1920.

[52] K. Smet, S. Jost-Boissard, W. R. Ryckaert, *et al.*, "Validation of a Colour Rendering Index Based on Memory Colours," in *Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2010, pp. 136–142.

[53] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "A Memory Colour Quality Metric for White Light Sources," *Energy and Buildings*, vol. 49, p. 216, 2012.

[54] K. Smet and P. Hanselaer, "Memory and Preferred Colours and the Colour Rendition of White Light Sources," *Lighting Research & Technology*, vol. 48, no. 4, p. 393, 2016.

[154] S. Jost-Boissard, P. Avouac, and M. Fontoynont, "Assessing the Colour Quality of LED Sources: Naturalness, Attractiveness, Colourfulness and Colour Difference," *Lighting Research & Technology*, vol. 47, no. 7, pp. 769–794, 2015.

[195] S. Jost-Boissard, M. Fontoynont, and J. Blanc-Gonnet, "Perceived Lighting Quality of LED Sources for the Presentation of Fruit and Vegetables," *Journal of Modern Optics*, vol. 56, no. 13, pp. 1420–1432, 2009.

[196] M. Royer, A. Wilkerson, M. Wei, *et al.*, "Human Perceptions of Colour Rendition Vary with Average Fidelity, Average Gamut, and Gamut Shape," *Lighting Research & Technology*, vol. 49, no. 8, pp. 966–991, 2017.

[219] H. Wang, G. Cui, M. R. Luo, *et al.*, "Evaluation of Colour-Difference Formulae for Different Colour-Difference Magnitudes," *Color Research & Application*, vol. 37, no. 5, pp. 316–325, 2012.

[220] M. Melgosa, P. A. García, L. Gómez-Robledo, *et al.*, "Notes on the Application of the Standardized Residual Sum of Squares Index for the Assessment of Intra- and Inter-observer Variability in Color-Difference Experiments," *Journal of the Optical Society of America A*, vol. 28, no. 5, pp. 949–953, 2011.

[281] C. Spearman, "The Proof and Measurement of Association between two Things," *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[282] W. D Wayne, "Spearman Rank Correlation Coefficient," in *Applied Nonparametric Statistics*. Boston, MA, USA: PWS-Kent, 1990, pp. 358–365.

[283] A. Rohatgi, *WebPlotDigitizer 4.0*, `https://automeris.io/WebPlotDigitizer/`, [Online; accessed 28-March-2018], 2017.

[284] M. Wei, K. W. Houser, G. R. Allen, *et al.*, "Color Preference under LEDs with Diminished Yellow Emission," *Journal of the Illuminating Engineering Society*, vol. 10, no. 3, pp. 119–131, 2014.

[285] F. Szabó, R. Kéri, J. Schanda, *et al.*, "A Study of Preferred Colour Rendering of Light Sources: Home Lighting," *Lighting Research & Technology*, vol. 48, no. 2, pp. 103–125, 2016.

[286] J. VanRie, "The Effect of the Spectral Composition of a Light Source on the Visual Appreciation of a Composite Object Set," in *Technical Report to the User Committee of the IWT-TETRA Project (80163), Appendix* 4, Diepenbeek, Belgium, 2009.

[287] PennState, Eberly College of Science, *STAT 503 Design of Experiments: The Latin Square Design*, `https://onlinecourses.science.psu.edu/stat503/node/21`, [Online; accessed 27-March-2018].

[288] Y. Imai, T. Kotani, and T. Fuchida, "A Study of Colour Rendering Properties based on Colour Preference in Adaptation to LED Lighting," in *Proceedings of the CIE 2012 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2012, pp. 369–374.

[289] Y. Imai, T. Kotani, and T. Fuchida, "A Study of Colour Rendering Properties based on Colour Preference of Objects in Adaptation to LED Lighting," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 62–67.

[290] S. Jost and M. Fontoynont, "Colour Rendering of Face Complexion and Hair under LED Sources," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 53–61.

[291] A. Tsukitani, "Optimization of Colour Quality for Landscape Lighting based on Feeling of Contrast Index," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 68–71.

[292] Y. Lin, J. He, A. Tsukitani, *et al.*, "Colour Quality Evaluation of Natural Objects based on the Feeling of Contrast Index," *Lighting Research & Technology*, vol. 48, no. 3, pp. 323–339, 2016.

[293] K. W. Houser, D. K. Tiller, and X. Hu, "Tuning the Fluorescent Spectrum for the Trichromatic Visual Response: A Pilot Study," *Journal of the Illuminating Engineering Society*, vol. 1, no. 1, pp. 7–23, 2005.

[294] Z. Huang, Q. Liu, S. Westland, *et al.*, "Light Dominates Colour Preference when Correlated Colour Temperature Differs," *Lighting Research & Technology*, 2017, [online-first].

[295] N. Narendran and L. Deng, "Color Rendering Properties of LED Light Sources," *Proceedings of SPIE*, vol. 4776, pp. 61–67, 2002.

[296] Q. Wang, H. Xu, F. Zhang, *et al.*, "Influence of Color Temperature on Comfort and Preference for LED Indoor Lighting," *Optik*, vol. 129, pp. 21–29, 2017.

[297] Q. Liu, Z. Huang, K. Xiao, *et al.*, "Gamut Volume Index: A Color Preference Metric based on Meta-analysis and Optimized Color Samples," *Optics Express*, vol. 25, no. 14, pp. 16 378–16 391, 2017.

[298] J. E. Hunter and F. L. Schmidt, *Methods of Meta-analysis: Correcting Error and Bias in Research findings*, 2nd. Thousand Oaks, CA, USA: SAGE Publications, 2004.

[299] J. Sánchez-Meca and F. Marín-Martínez, "Confidence Intervals for the Overall Effect Size in Random-Effects Meta-Analysis," *Psychological Methods*, vol. 13, no. 1, pp. 31–48, 2008.

[300] K. Sidik and J. N. Jonkman, "A Comparison of Heterogeneity Variance Estimators in Combining Results of Studies," *Statistics in Medicine*, vol. 26, no. 9, pp. 1964–1981, 2007.

[301] W. Viechtbauer, "Bias and Efficiency of Meta-analytic Variance Estimators in the Random-Effects Model," *Journal of Educational and Behavioral Statistics*, vol. 30, no. 3, pp. 261–293, 2005.

[302] G. Y. Zou, "Toward Using Confidence Intervals to Compare Correlations," *Psychological Methods*, vol. 12, no. 4, pp. 399–413, 2007.

[303] R. A. Fisher, "Frequency Distribution of the Values of the Correlation Coefficient in Samples of an Indefinitely Large Population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[304] R. A. Fisher, "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," *Metron*, vol. 1, pp. 3–32, 1921.

[305] Z. Ugray, L. Lasdon, J. Plummer, *et al.*, "Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization," *INFORMS Journal on Computing*, vol. 19, no. 3, pp. 328–340, 2007.

[306] P. Csuti and J. Schanda, "Colour Matching Experiments with RGB-LEDs," *Color Research & Application*, vol. 33, no. 2, pp. 108–112, 2008.

[307] P. Csuti and J. Schanda, "A Better Description of Metameric Experience of LED Clusters," *Light & Engineering*, vol. 18, no. 1, pp. 44–50, 2010.

[308] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 1: Experiments in a Room," *Lighting Research & Technology*, vol. 49, no. 6, pp. 697–713, 2017.

[309] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 2: Experiments in a Viewing Booth and Analysis of the Combined Dataset," *Lighting Research & Technology*, vol. 49, no. 6, pp. 714–726, 2017.

[310] T. Q. Khanh and P. Bodrogi, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 3: Experiments with Makeup Products and Analysis of the Complete Warm White Dataset," *Lighting Research & Technology*, vol. 50, no. 2, pp. 218–236, 2018.

[311] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 4: Experiments with Still Life Arrangements at Different Correlated Colour Temperatures," *Lighting Research & Technology*, 2017, [online-first].

[312] T. Q. Khanh, P. Bodrogi, X. Guo, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 5: A Colour Preference Experiment at 2000 lx in a Real Room," *Lighting Research & Technology*, 2017, [online-first].

[313] E. Schrödinger, *BrainyQuote.com: Quotes of Erwin Schrödinger*, `https://www.brainyquote.com/quotes/erwin_schrodinger_304801`, [Online; accessed 30-May-2018].

Part 6

CONCLUSION AND OUTLOOK

"What we observe as material bodies and forces are nothing but shapes and variations in the structure of space"

– Erwin Schrödinger [313]

6

CONCLUSION AND OUTLOOK

With their inherent ability of serving as an internal reference, memory colors have proven to provide a powerful concept for the evaluation of the color rendering properties of white light sources in terms of predicting visual appreciation. However, as discussed in this thesis, some major drawbacks could be identified in the principal design of existing memory-based or memory-related color quality metrics, of which the most severe is most likely that none of these metrics were devised under realistic adaptation and viewing conditions. For this reason, the first of this thesis' main goals was to establish a new experimental approach trying to overcome the reported shortcomings of previous studies on the assessment of memory colors.

The most prominent feature of this new approach, besides considering real familiar test objects, was the inclusion of realistic viewing and adaption conditions. Hence, an improved color appearance rating experiment was conducted by the author intended to conceive the characteristics of the memory colors of a set of familiar test objects by providing a more realistic contextual scenery. The respective color appearance ratings were performed by two different cultural observer groups (Chinese vs. German observers) which were exposed to two different adaptation conditions (3200 K vs. 5600 K, both at 2000 lx) in order to investigate both the impact of the adapted white point and the observers' cultural background on the memory color assessments, which was the second goal of this thesis. With the corresponding dependencies being eventually known, the third goal was finally to derive an updated and improved version of a memory-based color quality metric meant for giving a well-correlating measure for the visual appreciation of a light source.

With respect to the first and second goal to be addressed by the present work, Chapter 4 did not just give a detailed overview of the improved experimental setup, it also provided a thorough descriptive and statistical analysis of the color appearance ratings collected for each combination of the adapted white point and the observers' cultural background. In contrast to previous studies, the experiments, for the first time, were performed under realistic viewing and adaptation conditions in a real-sized furnished room providing an immersive and contextual viewing experience, which basically implied the possibility of drawing more universally valid conclusions from the current experimental data than from the results of previous experiments on the same topic. Besides making use of a better suited CIECAM02-based color space for the evaluation of the color appearance rating data including the sophisticated CAT02 transform to properly model the observers' state of chromatic adaptation, the focus was on a less arbitrary and more comprehensible test object selection, which could be achieved by a specifically designed online survey conducted in advance of preparing the new experiments. Especially the lack of a well-founded object selection was considered as a strong point of criticism of past studies. In Fig. 6.1, the corresponding CIECAM02-UCS chromaticities of the twelve test samples adopted in the current thesis are shown as perceived under reference illuminant D56 and compared to the nine colored test samples used in the experiments conducted by Smet *et al.* [40]. As can be seen, the newly selected test objects additionally provide an improved hue coverage.

In particular, the reddish part of the hue circle, which was found to play a more important role than other hues for modeling human preference [196], is much better represented by the newly selected test objects even in the more saturated areas. Furthermore, with the exception of orange and Smurf®, where the latter, as stated previously, must be considered as a very

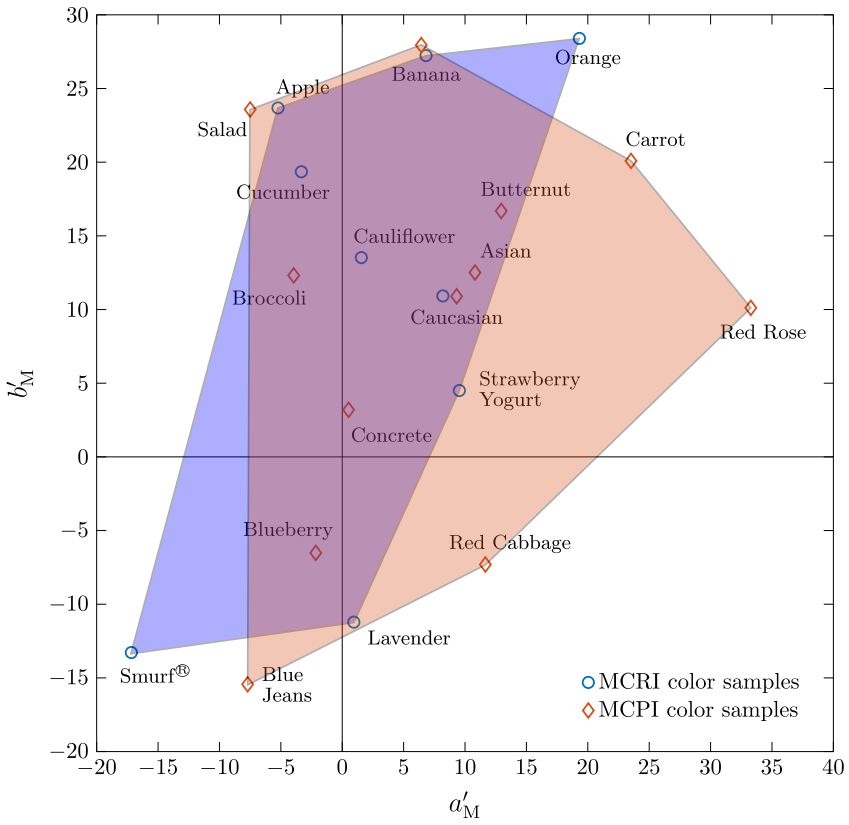


Figure 6.1 – Comparison of the CIECAM02-UCS chromaticities of the twelve MCPI test samples adopted in the current thesis and the chromaticities of the nine colored test samples used by Smet *et al.* [40] to define the MCRI. Reference illuminant D56 was assumed for the calculations.

arguable choice for this kind of experiments, all test objects adopted here for defining the new memory-based color quality metric show considerably higher degrees of saturation. Following the argumentation given in Refs. [19, 194], this is basically considered to enhance the metric's predictive performance especially for peaked LED spectra and, therefore, represents a very important feature of the new metric definition.

With all these improvements in the experimental design, it was of great interest whether or not the general tendencies of memory colors derived from various other studies [25, 31, 33–35, 37–41] could also be observed here. Based on the analysis of the color appearance rating data for the set of familiar test objects assessed by the two different cultural observer groups exposed to the two different adaptation conditions, it could clearly be justified by the obtained results that, despite the new experimental setup, the corresponding memory colors in comparison to their appearance under reference illumination still tend to be shifted towards higher chroma as well as towards their typical hues, which basically is in accordance to the findings reported in previous studies. With each combination of the cultural background of the observers and the light situation they are adapted to showing more or less the same tendencies regarding the memory-induced chromaticity shifts, the observed trends derived from the present color appearance rating experiments were additionally concluded to be an inherent feature of human color perception rather than being a matter of cultural peculiarity or adapted white point.

Despite showing similar characteristics, a noteworthy statistical effect on the assessment of memory colors had yet to be reported for both the changing adaptation conditions and the inter-cultural variations between Chinese and German observers. In particular, it was inferred from the outcome of the applied extra sum-of-squares F-tests that the variability

in the color appearance ratings of the familiar test objects between the individual observers of the same cultural background was slightly smaller than the variability observed for the ratings between the different cultural observer groups so that, even though the average effect size was quite small, the potential impact of the observers' cultural background as well as the more pronounced impact of the adaptation conditions, which showed a much larger effect size, were both found to be non-negligible. From the analysis of the respective effect sizes, an intrinsic hierarchy for the importance of these two different impact factors could further be derived. On average, the effect size of the inter-cultural variations in the memory color assessments were shown to be approximately 40 % smaller than the one corresponding to the impact of the adapted white point. Nevertheless, as argued above, both impact factors were concluded to be large enough for not being simply neglected for the intended creation of an updated memory-based color quality metric.

The development of such a metric for the evaluation of the color rendering properties of white light sources in terms of visual appreciation as the third goal of the current thesis was mainly pursued in Chapter 5. This newly proposed color quality metric denoted as memory color preference index MCPI was based on the evaluation of the degree of similarity between the color appearance of certain familiar test objects rendered by an arbitrary test light source and their respective memory color representations. The degree of similarity was assessed by using a set of Gaussian similarity distribution functions fitted to the results of the color appearance rating experiments of Chapter 4. The key features of this new proposal were the adoption of the perceptionally uniform CIECAM02-UCS as the working color space, the implementation of a CCT-based decision algorithm to choose a suitable set of similarity functions better approximating the impact of the adapted white point on the memory color assessments, and the introduction of some additional weighting factors allowing i) to model the varying importance of certain test colors in the evaluation of light sources with respect to color preference and ii) to counterbalance the metric errors in the memory color assessments introduced by chromatic adaptation. Moreover, three separate MCPI versions, i.e., a Chinese, a German, and a global one, where the latter was derived from the pooled Chinese and German observer data, were constructed in order to account for the potential inter-cultural differences in the evaluation of the perceived color quality of white light sources based on memory colors.

In order to subsequently validate the superiority of this new proposal over existing methods, a meta-correlation analysis comprising the experimental data of a comprehensive selection of different visual studies investigating color preference ratings of human observers in both metameric and multi-CCT lighting scenarios was performed on a representative variety of different color quality metrics. Based on this analysis, it could be shown that the newly proposed MCPI metric significantly outperformed all alternative approaches exhibiting a much better performance in terms of predicting visual appreciation of light sources. In this context, a thorough discussion for the reasons of the rather bad performance or even the failing of the alternatives was provided. Furthermore, no or only barely significant differences between the global and the cultural-specific MCPI definitions could be found with respect to their predictive properties indicating that the former is a sufficiently good approximation to be used as a single, universally valid memory-based color quality metric which performs significantly better than the alternatives considered in this work.

With the newly proposed MCPI offering excellent predictive performance, future research intentions will focus on the application of this specific color quality metric for the optimization of state-of-the-art LED light sources with the aim of achieving high user acceptability and visual appreciation in a broad variety of different applications. In this regard, it should

be investigated whether a variation in the visual context also demands an adjustment of the MCPI weighting parameters. Similar to the observed differences when comparing metameric and multi-CCT lighting scenarios, it cannot be precluded that due to context-induced cognitive processes a different lighting environment (e.g., office vs. home lighting) would also require a modified weighting of the individual memory color samples used for constructing the MCPI metric. Hence, the goal of future research on this topic would be to define – if required – for each possible field of application a context-specific version of the MCPI, which provides a reliable and easy-to-implement optimization tool for the industry capable of finally replacing the CIE R_a metric in cases where visual appreciation and color preference are more important than color fidelity.

A further research question, that has been ignored so far, concerns the comparison of the distribution of the absolute color appearance ratings given by the individual observers for a specific object chromaticity. When comparing for example the maximum values of the fitted Gaussian similarity distribution functions for each test object assessed by the German observer group at 5600 K and 3200 K ambient illumination (see Figs. 4.17 and 4.24, respectively), it seems as if the latter adaptation condition yields a slightly smaller maximum rating than the former for some of the test objects indicating a somewhat stronger preference of the observers for the most likely memory color of the familiar test objects when being adapted to a more cool-white illumination. For the Chinese observers, a similar finding can be reported. Hence, the question arises whether or not these differences in the observers' absolute ratings between different adaptation conditions are significant and, if so, how one can use this additional information to further enhance the performance of the MCPI metric. It might for example be expected that especially for multi-CCT scenarios, where the change in the white point of adaption evidently has a large impact on the observers' color preference ratings, an additional improvement could probably be achieved in case that the deviations in the absolute color appearance ratings were adequately reflected in the MCPI definition. Furthermore, one might ask if such an effect could also be observed between the ratings of different cultural observer groups but for the same white point of the ambient illumination.

Last but not least and also with regard to the previous paragraph, an extension of the experimental approach presented in this thesis to a greater selection of different adaptation conditions and cultural observer groups would certainly be a great benefit for the refinement of the MCPI color quality metric but would also allow for a more comprehensive analysis of the characteristics of memory colors. If possible, future experiments should be conducted entirely in the native environment of the respective cultural observer group in order to avoid unwanted bias caused by an advancing familiarization with another culture. For the Chinese observer group tested in the current thesis such an effect could not be precluded completely, even though care was taken to invite only those Chinese to participate in the experiments who had been living in Germany for only a quite short period of time so that their color appearance ratings could still be assumed as "pure" Chinese. Of course, it would be better to (re-)perform the experiments in a genuine setup in the observers' country of origin, which however might be very difficult to organize but should anyhow be in the focus of future research on this topic.

Part A

APPENDIX

A

APPENDIX

A.1 COLORIMETRIC DATA OF MCPI TEST SAMPLES

For each of the 12 familiar test objects selected for performing the new experiments on the assessment of memory colors, the corresponding colorimetric properties were calculated for the two different reference illuminants, i.e., the Planckian radiator at 3200 K and the CIE D56 standardized daylight phase, adopted in this thesis. In Table A.1, the resulting CIECAM02-UCS color coordinates J', a'_M, and b'_M as well as the CIECAM02 correlates of chroma C and hue h are summarized providing a better overview.

Table A.1 – Colorimetric data of the MCPI test samples as perceived under reference illumination represented by the 3200 K Planckian radiator and the CIE D56 standard, respectively.

	3200 K Planckian radiator					CIE D56 reference illuminant				
Test Object	J'	a'_M	b'_M	h in °	C	J'	a'_M	b'_M	h in °	C
Asian Skin	66.86	13.94	13.01	43.02	25.30	65.72	10.81	12.51	49.18	21.28
Banana	82.55	6.20	28.21	77.60	43.31	81.31	6.44	27.94	77.02	42.88
Blueberry	38.42	−1.36	−5.37	255.77	6.25	38.55	−2.16	−6.52	251.70	7.87
Blue Jeans	35.67	−7.54	−14.89	243.14	21.51	36.25	−7.71	−15.43	243.47	22.39
Broccoli	42.21	−4.98	12.40	111.88	16.55	42.28	−3.95	12.32	107.79	15.94
Butternut Squash	64.89	13.19	17.77	53.43	30.49	63.41	12.95	16.69	52.19	28.75
Carrot	67.90	23.69	21.66	42.44	50.12	65.23	23.52	20.08	40.50	47.59
Caucasian Skin	72.57	12.73	11.38	41.78	22.11	71.62	9.32	10.90	49.47	17.97
Concrete Flowerpot	71.81	1.50	4.15	70.13	4.92	71.72	0.52	3.19	80.82	3.56
Green Salad	55.95	−9.83	23.08	113.08	35.86	56.07	−7.50	23.58	107.64	35.22
Red Cabbage	34.06	15.33	−4.55	343.48	20.45	33.27	11.66	−7.30	327.93	17.12
Red Rose	31.63	35.70	12.42	19.18	63.53	28.55	33.27	10.11	16.91	56.22

A.2 CONTOUR LINE PLOTS OF CHINESE VS. GERMAN OBSERVERS

In order to draw valid conclusions about the effect size and the significance of the impact of the observers' cultural background on the color appearance ratings of familiar test objects, further statistical analysis based on multivariate methods must be performed. Like for the analysis of the impact of different adaptation conditions on the memory color assessments, contour line representations of the fitted similarity distribution functions can be used to visually compare the average ratings of the Chinese observers with those obtained for their German counterparts and are illustrated in Figs. A.1 to A.4 on the following pages.

As can be seen, somewhat larger deviations in the ellipses' shape, size, and orientation between the similarity distributions of both cultural observer groups are found for the adaptation conditions at 3200 K than for those at 5600 K, which basically is in accordance to the results of Box's M-test reported in Table 4.25 of Sec. 4.4.3. As expected from the respective contour line plots, significant differences in the covariance matrices of the similarity distributions can be found for the test objects of broccoli, butternut, concrete, green salad, red cabbage, and red rose for the 3200 K ambient illumination, whereas in case of the 5600 K adaptation conditions significant differences are concluded only for the test objects of broccoli and red rose. On the other hand, regarding the for some test objects barely noticeable shifts in the ellipses' centroid locations, Hotelling's t^2-test revealed no significance at all so that for both adaptation conditions the null hypothesis of equal sample mean vectors between the modeled color appearance ratings of Chinese and German observers cannot be rejected. Hence, as argued in this thesis, even though significant deviations are observed for the various ellipse parameters, the inter-cultural effect on the assessment of memory colors must be inferred to be quite small, which is reflected in the relatively small effect sizes calculated for the outcome of Hotelling's t^2-test.

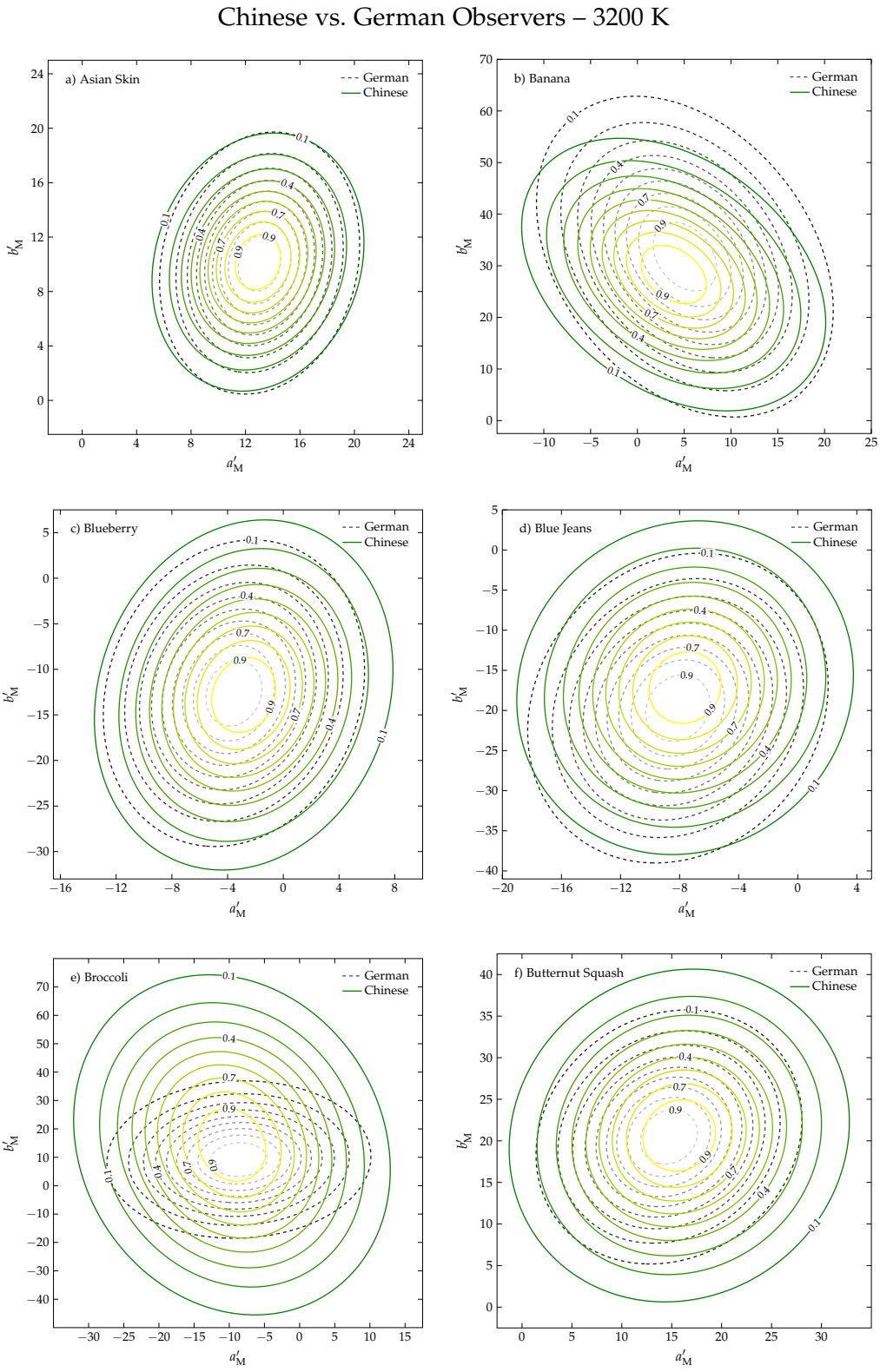


Figure A.1 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the Chinese (green-to-yellow colormap, solid line) and German (dark-to-light-gray colormap, dashed line) observers for the test objects of a) asian skin, b) banana, c) blueberry, d) blue jeans, e) broccoli, and f) butternut squash assessed under 3200 K ambient illumination.

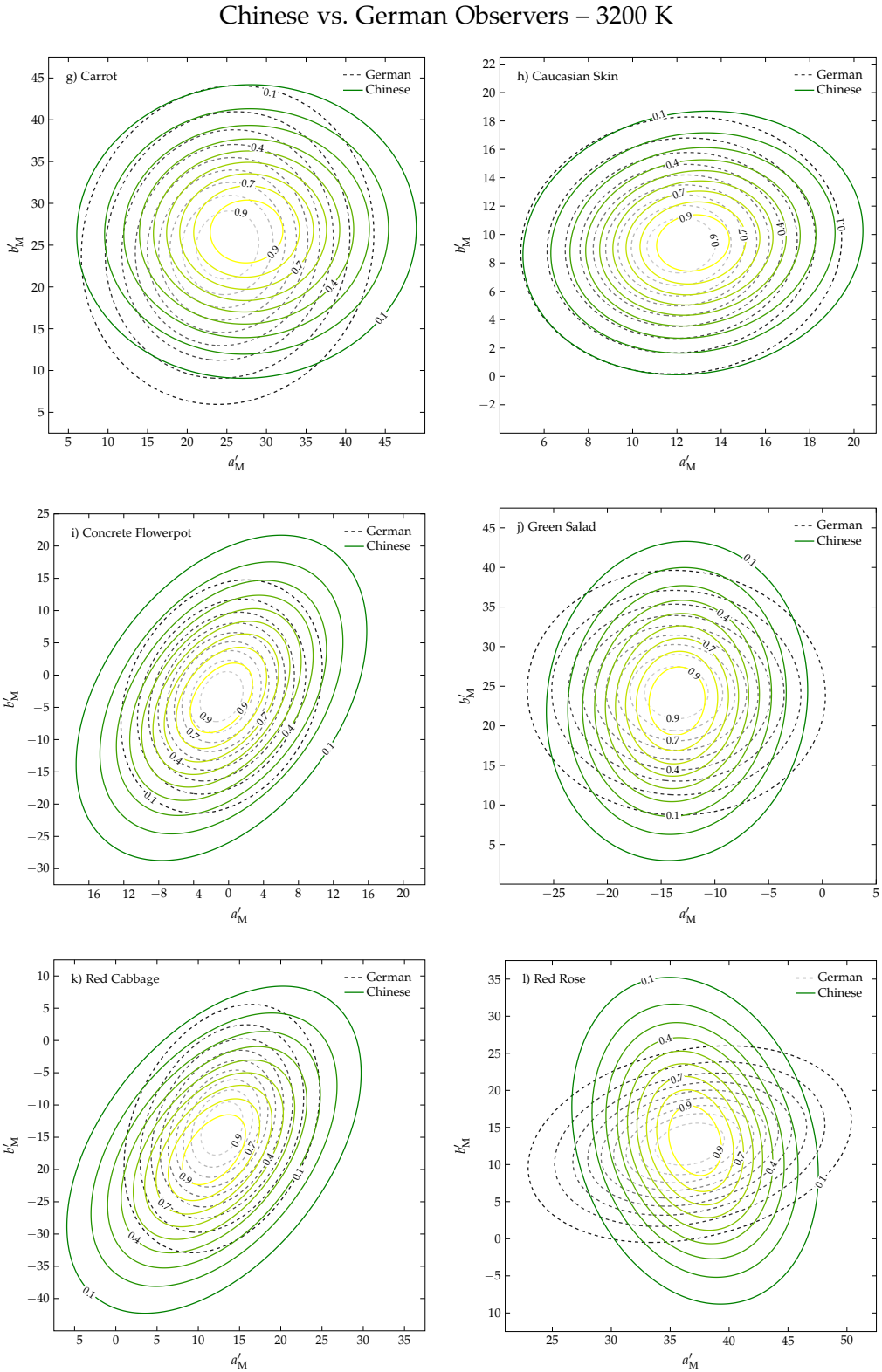


Figure A.2 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the Chinese (green-to-yellow colormap, solid line) and German (dark-to-light-gray colormap, dashed line) observers for the test objects of g) carrot, h) Caucasian skin, i) concrete flowerpot, j) green salad, k) red cabbage, and l) red rose assessed under 3200 K ambient illumination.

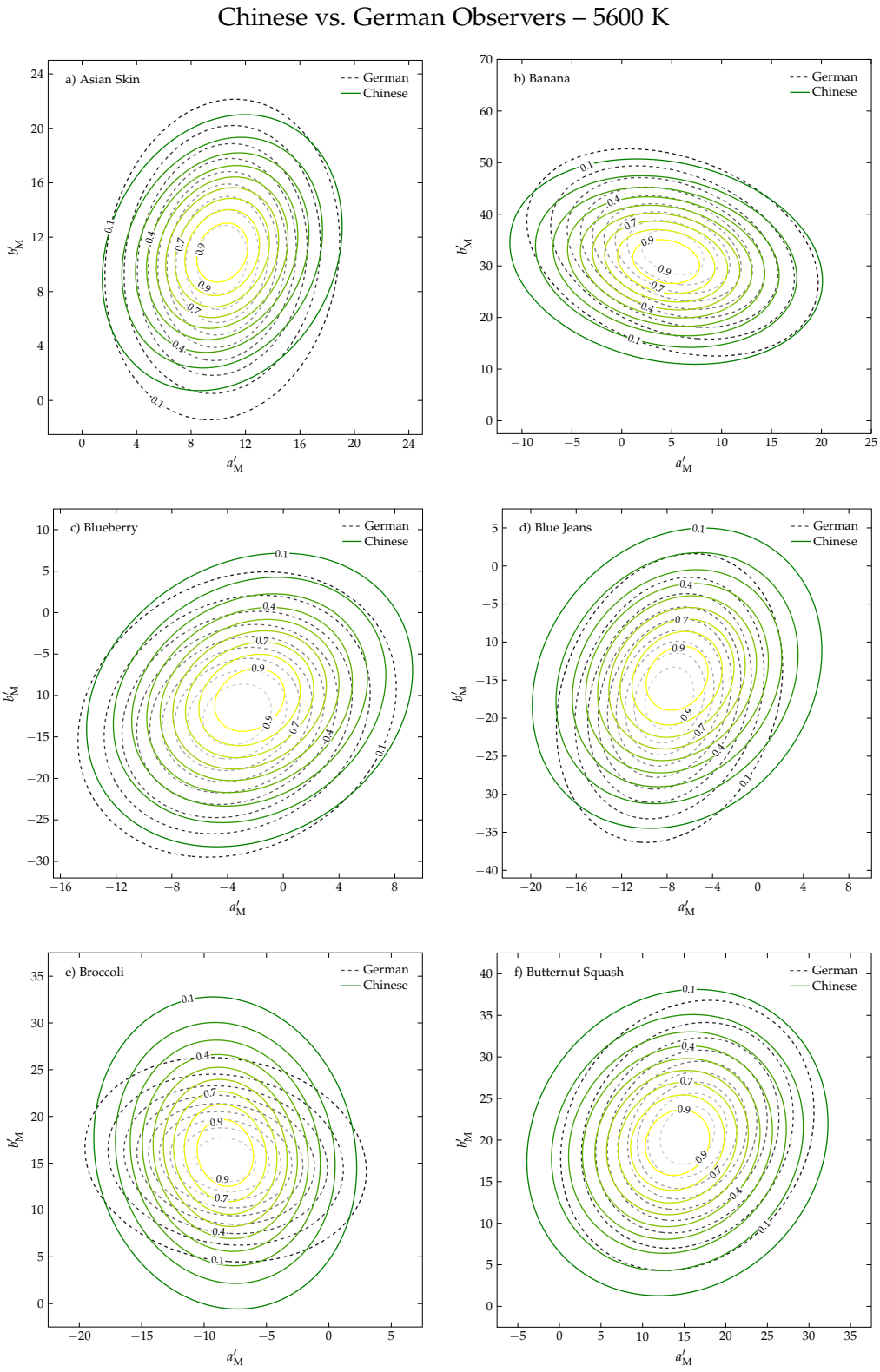


Figure A.3 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the Chinese (green-to-yellow colormap, solid line) and German (dark-to-light-gray colormap, dashed line) observers for the test objects of a) asian skin, b) banana, c) blueberry, d) blue jeans, e) broccoli, and f) butternut squash assessed under 5600 K ambient illumination.

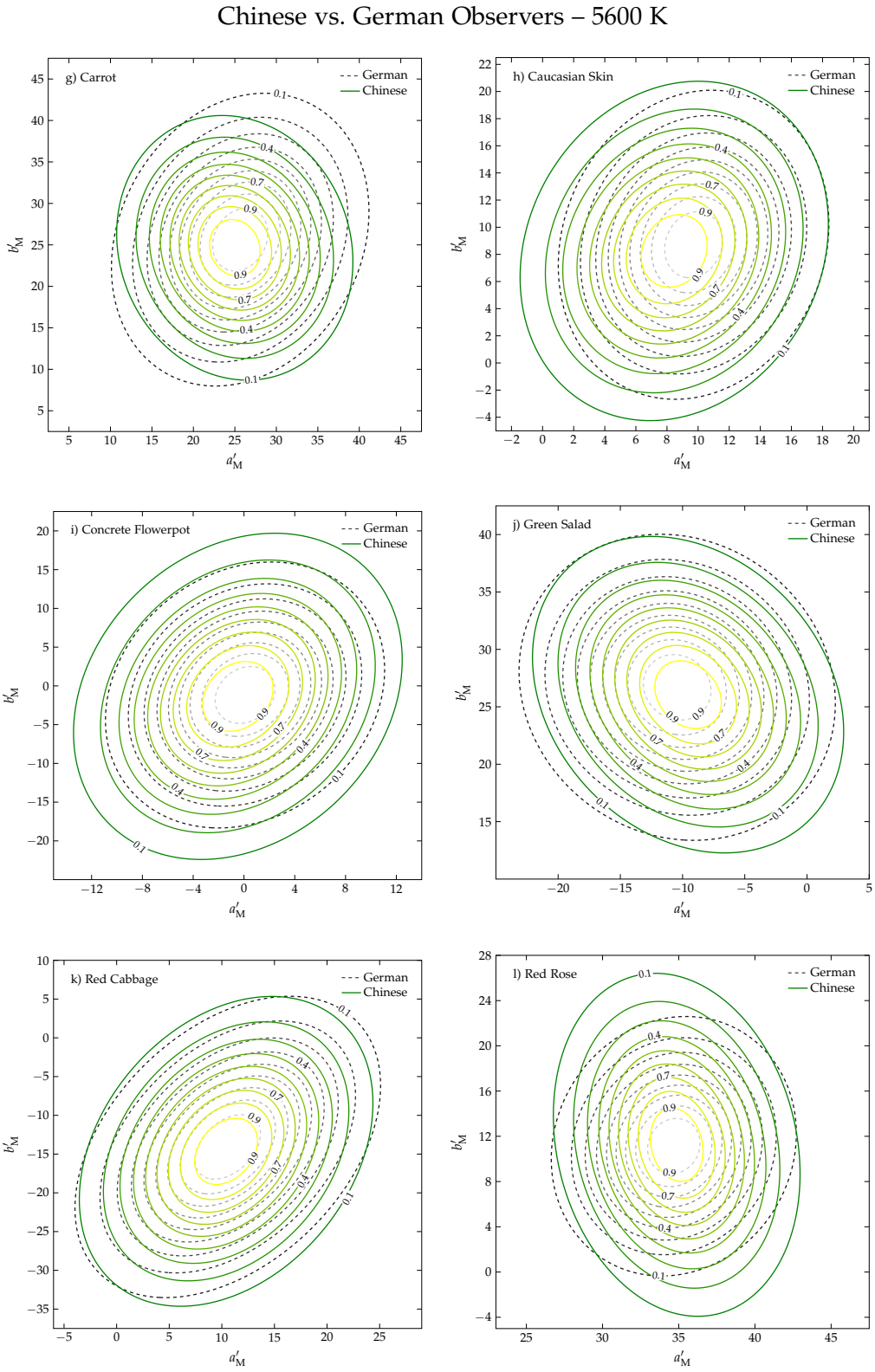


Figure A.4 – Contour line plots of the fitted, normalized similarity distribution functions obtained for the Chinese (green-to-yellow colormap, solid line) and German (dark-to-light-gray colormap, dashed line) observers for the test objects of g) carrot, h) Caucasian skin, i) concrete flowerpot, j) green salad, k) red cabbage, and l) red rose assessed under 5600 K ambient illumination.

BIBLIOGRAPHY

[1] A. Hutchins, *Playing with Monsters*. Spokane, WA, USA: Self-published, Kindle Edition, 2016.

[2] A. Einstein, *Out of My Later Years: The Scientist, Philosopher, and Man Portrayed Through His Own Words*. New York, NY, USA: Castle Books, 2005.

[3] J. Levy, *Really Useful: The Origins of Everyday Things*. Firefly Books, Berlin, Germany, 2002.

[4] T. A. Edison, "Electric-Lamp," Patent US 223898 A, Jan. 1880.

[5] L. H. Latimer, "Process of Manufactoring Carbons," Patent US 252386 A, Jan. 1882.

[6] C. B. Davenport, "The Nernst Lamp," *Science*, vol. 8, no. 203, p. 689, 1898.

[7] L. R. Ingersoli, "On the Radiant Efficiency of the Nernst Lamp," *Physical Review*, vol. 17, no. 5, p. 371, 1903.

[8] T. Welker, "Recent Developments on Phosphors for Fluorescent Lamps and Cathode-Ray Tubes," *Journal of Luminescence*, vol. 48–49, no. 1, p. 49, 1991.

[9] A. M. Srivastava and T. Sommerer, "Fluorescent Lamp Phosphors," *Interface*, vol. 7, no. 2, p. 28, 1998.

[10] F. Grum, "Artificial Light Sources for Simulating Natural Daylight and Skylight," *Applied Optics*, vol. 7, no. 1, p. 183, 1968.

[11] S. J. Dain, "Daylight Simulators and Colour Vision Tests," *Ophthalmic and Physiological Optics*, vol. 18, no. 6, p. 540, 1998.

[12] H. Xu, M. R. Luo, and B. Rigg, "Evaluation of Daylight Simulators Part I: Colorimetric and Spectral Variations," *Coloration Technology*, vol. 119, no. 2, p. 59, 2003.

[13] D. Nickerson, "Measurement and Specification of Color Rendition Properties of Light Sources," *Illuminating Engineering*, vol. 53, no. 2, p. 77, 1958.

[14] J. L. Ouweltjes, "The Specification of Colour Rendering Properties of Fluorescent Lamps," *Die Farbe*, vol. 9, p. 207, 1960.

[15] D. Nickerson, "Light Sources and Colour rendering," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 57, 1960.

[16] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.1, 1965.

[17] D. Nickerson and C. W. Jerome, "Color Rendering of Light Sources: CIE Method of Specification and its Application," *Illuminating Engineering*, vol. 60, no. 4, p. 262, 1965.

[18] S. Jost-Boissard and M. Fontoynont, "Optimization of LED-based Light Blendings for Object Presentation," *Color Research & Application*, vol. 34, no. 4, p. 310, 2009.

[19] W. Davis and Y. Ohno, "Color quality scale," *Optical Engineering*, vol. 49, no. 3, p. 033 602, 2010.

[20] M. S. Islam, R. Dangol, M. Hyvärinen, *et al.*, "Investigation of User Preferences for LED Lighting in Terms of Light Spectrum," *Lighting Research & Technology*, vol. 45, no. 6, pp. 641–665, 2013.

[21] R. Dangol, M. Islam, M. Hyvärinen, *et al.*, "Subjective Preferences and Colour Quality Metrics of LED Light Sources," *Lighting Research & Technology*, vol. 45, p. 666, 2013.

[22] R. R. Baniya, R. Dangol, P. Bhusal, *et al.*, "User-acceptance Studies for Simplified Light-Emitting Diode Spectra," *Lighting Research & Technology*, vol. 47, p. 177, 2015.

[23] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Memory Colours and Colour Quality Evaluation of Conventional and Solid-State Lamps," *Optics Express*, vol. 18, no. 25, pp. 26 229–26 244, 2010.

[24] G. B. Buck and H. C. Froelich, "Color Characteristics of Human Complexions," *Illuminating Engineering*, vol. 43, no. 1, p. 27, 1948.

[25] C. L. Sanders, "Color Preferences for Natural Objects," *Illuminating Engineering*, vol. 54, p. 452, 1959.

[26] C. L. Sanders, "Assessment of Color Rendition under an Illuminant Using Color Tolerances for Natural Objects," *Illuminating Engineering*, vol. 54, p. 640, 1959.

[27] D. B. Judd, "A Flattery Index for Artificial Illuminants," *Illuminating Engineering*, vol. 62, p. 593, 1967.

[28] C. W. Jerome, "Flattery vs. Color Rendition," *Journal of the Illuminating Engineering Society*, vol. 1, no. 3, p. 208, 1972.

[29] C. W. Jerome, "The Flattery Index," *Journal of the Illuminating Engineering Society*, vol. 2, no. 4, p. 351, 1973.

[30] W. A. Thornton, "A Validation of the Color-Preference Index," *Journal of the Illuminating Engineering Society*, vol. 4, no. 1, p. 48, 1974.

[31] P. Siple and R. M. Springer, "Memory and Preference for the Colors of Objects," *Perception & Psychophysics*, vol. 34, no. 4, p. 363, 1983.

[32] E. Hering, *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer Verlag, 1920.

[33] C. J. Bartleson, "Color in Memory in Relation to Photographic Reproduction," *Photographic Science and Engineering*, vol. 5, no. 6, p. 327, 1961.

[34] S. M. Newhall, R. W. Burnham, and J. R. Clark, "Comparison of Successive with Simultaneous Color Matching," *Journal of the Optical Society of America*, vol. 47, no. 1, p. 43, 1957.

[35] C. J. Bartleson, "Memory Colors of Familiar Objects," *Journal of the Optical Society of America*, vol. 50, no. 1, p. 73, 1960.

[36] C. D. Hendley and S. Hecht, "The Colors of Natural Objects and Terrains, and their Relation to Visual Color Deficiency," *Journal of the Optical Society of America*, vol. 39, no. 10, p. 870, 1949.

[37] J. Pérez-Carpinell, M. D. de Fez, R. Baldoví, *et al.*, "Familiar Objects and Memory Color," *Color Research & Application*, vol. 23, no. 6, pp. 416–427, 1998.

[38] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Representation of Memory Prototype for an Object Color," *Color Research & Application*, vol. 24, no. 6, pp. 393–410, 1999.

[39] P. Bodrogi and T. Tarczali, "Colour Memory for Various Sky, Skin, and Plant Colours: Effect of the Image context," *Color Research & Application*, vol. 26, no. 4, p. 278, 2001.

[40] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Colour Appearance Rating of Familiar Real Objects," *Color Research & Application*, vol. 36, no. 3, pp. 192–200, 2011.

[41] T. Tarczali, S. Du Park, P. Bodrogi, *et al.*, "Long-term Memory Colors of Korean and Hungarian Observers," *Color Research & Application*, vol. 31, no. 3, p. 176, 2006.

[42] A. C. Hurlbert and Y. Ling, "If It's a Banana, It must be Yellow: The Role of Memory Colors in Color Constancy," *Journal of Vision*, vol. 5, no. 8, p. 787, 2005.

[43] T. Hansen, M. Olkkonen, S. Walter, *et al.*, "Memory Modulates Color Appearance," *Nature Neuroscience*, vol. 9, no. 11, p. 1367, 2006.

[44] J. J. M. Granzier and K. R. Gegenfurtner, "Effects of Memory Color on Color Constancy for Unknown Colored Objects," *i-Perception*, vol. 3, no. 3, p. 190, 2012.

[45] P. Bodrogi and T. Tarczali, "Investigation of Colour Memory," in *Colour Image Science: Exploiting Digital Media*, L. W. MacDonald and M. R. Luo, Eds. Chichester: John Wiley & Sons Limited, 2002, pp. 23–48.

[46] H. Zeng and R. Luo, "Modelling Memory Color Region for Preference Color Reproduction," in *Proceedings of the SPIE 7528, Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*, San Jose, CA, USA: International Society for Optics and Photonics, 2010, p. 752 808.

[47] S. Xue, M. Tan, A. McNamara, *et al.*, "Exploring the Use of Memory Colors for Image Enhancement," in *Proceedings of the SPIE 9014, Human Vision and Electronic Imaging XIX*, San Francisco, CA, USA: International Society for Optics and Photonics, 2014, p. 901 411.

[48] C. Boust, H. Chahine, M. B. Chouikha, *et al.*, "Color Correction Judgements of Digital Images by Experts and Naive Observer," in *Proceedings of the PICS Conference 2003*, Rochester, NY, USA: Society for Imaging Science and Technology (IS&T), 2003, pp. 4–9.

[49] C. Boust, H. Brettel, F. Viénot, *et al.*, "Color Enhancement of Digital Images by Experts and Preference Judgements by Observers," *Journal of Imaging Science and Technology*, vol. 50, no. 1, pp. 1–11, 2006.

[50] C. Boust, F. Cittadini, M. B. Chouikha, *et al.*, "Does an Expert Use Memory Colors to Adjust Images?" In *Proceedings of the 12th Color and Imaging Conference: Color Science, Systems, and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 2004, pp. 347–353.

[51] S. N. Yendrikhovskij, F. J. J. Blommaert, and H. de Ridder, "Color Reproduction and the Naturalness Constraint," *Color Research & Application*, vol. 24, no. 1, pp. 52–67, 1999.

[52] K. Smet, S. Jost-Boissard, W. R. Ryckaert, *et al.*, "Validation of a Colour Rendering Index Based on Memory Colours," in *Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2010, pp. 136–142.

[53] K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "A Memory Colour Quality Metric for White Light Sources," *Energy and Buildings*, vol. 49, p. 216, 2012.

[54] K. Smet and P. Hanselaer, "Memory and Preferred Colours and the Colour Rendition of White Light Sources," *Lighting Research & Technology*, vol. 48, no. 4, p. 393, 2016.

[55] K. A. G. Smet, Y. Lin, B. V. Nagy, *et al.*, "Cross-Cultural Variation of Memory Colors of Familiar Objects," *Optics Express*, vol. 22, no. 26, p. 32 308, 2014.

[56] H. Matisse, *Quotes of Henri Matisse*, `https://quotefancy.com/quote/1315121/Henri-Matisse-Color-was-not-given-to-us-in-order-that-we-imitate-Nature-It-was-given-to`, [Online; accessed 22-May-2018].

[57] J. D. Mollon, "Color Vision," *Annual Review of Psychology*, vol. 33, pp. 41–85, 1982.

[58] J. D. Mollon and J. K. Bowmaker, "The Spatial Arrangement of Cones in the Primate Fovea," *Nature*, vol. 360, no. 6405, pp. 38–44, 1992.

[59] H. Hofer, J. Carroll, J. Neitz, *et al.*, "Organization of the Human Trichromatic Cone Mosaic," *Journal of Neuroscience*, vol. 25, no. 42, pp. 9669–9679, 2005.

[60] D. Williams, "Color and the Cone Mosaic," in *Proceedings of the 14th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2006, pp. 1–2.

[61] D. H. Brainard, "Color and the Cone Mosaic," *Annual Review of Vision Science*, vol. 1, pp. 519–546, 2015.

[62] A. Stockman and L. T. Sharpe, "The Spectral Sensitivities of the Middle- and Long-Wavelength-Sensitive Cones Derived from Measurements in Observers of known Genotype," *Vision Research*, vol. 40, no. 13, pp. 1711–1737, 2000.

[63] A. Stockman, L. T. Sharpe, and C. Fach, "The Spectral Sensitivity of the Human Short-wavelength Sensitive Cones Derived from Thresholds and Color Matches," *Vision Research*, vol. 39, no. 17, pp. 2901–2927, 1999.

[64] Commission Internationale de l'Éclairage, "Fundamental Chromaticity Diagram with Physiological Axes – Part 1," *CIE Technical Report 170-1:2006*, 2006.

[65] H. J. A. Dartnall, J. K. Bowmaker, and J. D. Mollon, "Human Visual Pigments: Microspectrophotometric Results from the Eyes of Seven Persons," *Proceedings of the Royal Society of London B*, vol. 220, no. 1218, pp. 115–130, 1983.

[66] J. K. Bowmaker and H. J. A. Dartnall, "Visual Pigments of Rods and Cones in a Human Retina," *Journal of Physiology*, vol. 298, no. 1, pp. 501–511, 1980.

[67] P. K. Brown and G. Wald, "Visual Pigments in Single Rods and Cones of the Human Retina," *Science*, vol. 144, no. 3614, pp. 45–46 & 51–52, 1964.

[68] R. W. G. Hunt and M. R. Pointer, *Measuring Colour*, 4th. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2011.

[69] J. Schanda, "CIE Colorimetry," in *Colorimetry: Understanding the CIE System*, J. Schanda, Ed., CIE Preprint Edition. Vienna: CIE Central Bureau, 2006, ch. 3, pp. 1–54.

[70] H. G. Grassmann, "Zur Theorie der Farbenmischung," *Annalen der Physik*, vol. 165, no. 5, pp. 69–84, 1853.

[71] W. D. Wright, "A Re-determination of the Trichromatic Coefficients of the Spectral Colours," *Transactions of the Optical Society*, vol. 30, no. 4, pp. 141–164, 1929.

[72] J. Guild, "The Colorimetric Properties of the Spectrum," *Philosophical Transactions of the Royal Society of London A*, vol. 230, pp. 149–187, 1931.

[73] Commission Internationale de l'Éclairage, *Colorimetry*, CIE Technical Report 15:2004, 3rd ed. 2004.

[74] Commission Internationale de l'Éclairage, "The Basis of Physical Photometry," *CIE Technical Report 18.2*, 1983.

[75] G. Wiszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd, Wiley Classics Library Edition. New York, NY, USA: John Wiley & Sons, Inc., 2000.

[76] D. Jameson and L. M. Hurvich, "Essay Concerning Color Constancy," *Annual Review of Psychology*, vol. 40, pp. 1–22, 1989.

[77] D. H. Foster, "Color Constancy," *Vision Research*, vol. 51, no. 7, pp. 674–700, 2011.

[78] A. Roorda and D. R. Williams, "The Arrangement of the Three Cone Classes in the Living Human Eye," *Nature*, vol. 397, no. 6719, pp. 520–522, 1999.

[79] J. Carroll, J. Neitz, and M. Neitz, "Estimates of L:M Cone Ratio from ERG Flicker Photometry and Genetics," *Journal of Vision*, vol. 2, no. 8, pp. 531–542, 2002.

[80] W. S. Stiles and J. Burch, "N.P.L. Colour-matching Investigation: Final Report," *Optica Acta*, vol. 6, no. 1, pp. 1–26, 1959.

[81] N. I. Speranskaya, "Determination of Spectrum Color Coordinates for 27 Normal Observers," *Optics and Spectroscopy*, vol. 7, pp. 424–428, 1959.

[82] D. L. MacAdam, "Visual Sensitivities to Color Differences in Daylight," *Journal of the Optical Society of America*, vol. 32, no. 5, pp. 247–274, 1942.

[83] D. L. MacAdam, "Specification of Small Chromaticity Differences," *Journal of the Optical Society of America*, vol. 33, no. 1, pp. 18–26, 1943.

[84] D. B. Judd, "A Maxwell Triangle Yielding Uniform Chromaticity Scales," *Journal of the Optical Society of America*, vol. 25, no. 1, pp. 24–35, 1935.

[85] D. B. Judd, "Estimation of Chromaticity Differences and Nearest Color Temperature on the Standard 1931 I. C. I. Colorimetric Coordinate System," *Journal of the Optical Society of America*, vol. 26, no. 11, pp. 421–426, 1936.

[86] D. L. MacAdam, "Projective Transformations of I. C. I. Color Specifications," *Journal of the Optical Society of America*, vol. 27, no. 8, pp. 294–299, 1937.

[87] Commission Internationale de l'Éclairage, "Technical Note: Brussels Session of the International Commission on Illumination," *Journal of the Optical Society of America*, vol. 50, no. 1, pp. 89–90, 1960.

[88] A. R. Robertson, "The CIE 1976 Color-Difference Formulae," *Color Research & Application*, vol. 2, no. 1, pp. 7–11, 1977.

[89] S. M. Jaekel, "Utility of Color-difference Formulas for Match-acceptability Decisions," *Applied Optics*, vol. 12, no. 6, pp. 1299–1316, 1973.

[90] M. R. Pointer, "A Comparison of the CIE 1976 Colour Spaces," *Color Research & Application*, vol. 6, no. 2, pp. 108–118, 1981.

[91] D. I. Morley, R. Munn, and F. W. Billmeyer, "Small and Moderate Colour Differences: II. The Morley Data," *Journal of the Society of Dyers and Colourists*, vol. 91, no. 7, pp. 229–242, 1975.

[92] F. J. J. Clarke, R. McDonald, and B. Rigg, "Modification to the JPC79 Colour–difference Formula," *Journal of the Society of Dyers and Colourists*, vol. 100, no. 4, pp. 128–132, 1984.

[93] Commission Internationale de l'Éclairage, "Improvement to Industrial Colour Difference Evaluation," *CIE Technical Report 142:2001*, 2001.

[94] M. R. Luo, G. Cui, and B. Rigg, "The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000," *Color Research & Application*, vol. 26, no. 5, pp. 340–350, 2001.

[95] G. Sharma, W. Wencheng, and E. N. Dalal, "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.

[96] E. Rohner and D. C. Rich, "Eine Angenähert Gleichförmige Metrik für Industrielle Farbtoleranzen von Körperfarben," *Die Farbe*, vol. 42, no. 4, pp. 207–220, 1996.

[97] M. R. Luo and C. Li, "CIECAM02 and Its Recent Developments," in *Advanced Color Image Processing and Analysis*, C. Fernandez-Maloigne, Ed. New York, USA: Springer Science+Business Media, 2013, pp. 19–58.

[98] M. Melgosa, R. Huertas, and R. S. Berns, "Relative Significance of the Terms in the CIEDE2000 and CIE94 Color-difference Formulas," *Journal of the Optical Society of America A*, vol. 21, no. 12, pp. 2269–2275, 2004.

[99] M. Melgosa, "Improvement of CMC upon CIEDE2000 for a New Experimental Dataset," *Color Research & Application*, vol. 31, no. 3, pp. 239–241, 2006.

[100] S. Shen and R. S. Berns, "Color-difference Formula Performance for Several Datasets of Small Color Differences Based on Visual Uncertainty," *Color Research & Application*, vol. 36, no. 1, pp. 15–26, 2011.

[101] G. Cui, M. R. Luo, B. Rigg, *et al.*, "Colour-Difference Evaluation Using CRT Colours. Part I: Data Gathering and Testing Colour Difference Formulae," *Color Research & Application*, vol. 26, no. 5, pp. 394–402, 2001.

[102] M. R. Luo, "Development of Colour-difference Formulae," *Review of Progress in Coloration and Related Topics*, vol. 32, no. 1, pp. 28–39, 2002.

[103] T. Seim and A. Valberg, "Towards a Uniform Color Space: A Better Formula to Describe the Munsell and OSA Color Scales," *Color Research & Application*, vol. 11, no. 1, pp. 11–24, 1986.

[104] Y. Nayatani, K. Takahama, and H. Sobagaki, "Prediction of Color Appearance under Various Adaptation Conditions," *Color Research & Application*, vol. 11, no. 1, pp. 62–71, 1986.

[105] Y. Nayatani, K. Hashimoto, K. Takahama, *et al.*, "Whiteness-blackness and Brightness Response in a Non-linear Color-appearance Model," *Color Research & Application*, vol. 12, no. 3, pp. 121–127, 1987.

[106] Y. Nayatani, K. Hashimoto, K. Takahama, *et al.*, "A Non-linear Color-appearance Model Using Estévez-Hunt-Pointer Primaries," *Color Research & Application*, vol. 12, no. 5, pp. 231–242, 1987.

[107] Y. Nayatani, K. Takahama, H. Sobagaki, *et al.*, "Color Appearance Model and Chromatic Adaptation Transform," *Color Research & Application*, vol. 15, no. 4, pp. 210–221, 1990.

[108] Y. Nayatani, H. Sobagaki, K. Hashimoto, *et al.*, "Field Trials of a Non-linear Color-appearance Model," *Color Research & Application*, vol. 22, no. 4, pp. 240–258, 1997.

[109] M. D. Fairchild and R. S. Berns, "Image Color-appearance specification through Extension of CIELAB," *Color Research & Application*, vol. 18, no. 3, pp. 178–190, 1993.

[110] M. D. Fairchild, "Refinement of the RLAB Color Space," *Color Research & Application*, vol. 21, no. 5, pp. 338–346, 1996.

[111] R. W. G. Hunt, "Perceptual Factors Affecting Colour Order Systems," *Color Research & Application*, vol. 10, no. 1, pp. 12–19, 1985.

[112] R. W. G. Hunt, "A Model for Colour Vision for Predicting Colour Appearance," *Color Research & Application*, vol. 7, no. 2, pp. 95–112, 1982.

[113] R. W. G. Hunt and M. R. Pointer, "A Colour-appearance Transform for the CIE 1931 Standard Colorimetric Observer," *Color Research & Application*, vol. 10, no. 3, pp. 165–179, 1985.

[114] R. W. G. Hunt, "A Visual Model for Predicting Colour Appearance under Various Viewing Conditions," *Color Research & Application*, vol. 12, no. 6, pp. 297–314, 1987.

[115] R. W. G. Hunt, "Hue Shifts in Unrelated and Related Colours," *Color Research & Application*, vol. 14, no. 5, pp. 235–239, 1989.

[116] R. W. G. Hunt, "Revised Colour-appearance Model for Related and Unrelated Colours," *Color Research & Application*, vol. 16, no. 3, pp. 146–165, 1991.

[117] R. W. G. Hunt, "An Improved Predictor of Colourfulness in a Model of Colour Vision," *Color Research & Application*, vol. 19, no. 1, pp. 23–26, 1994.

[118] R. W. G. Hunt and M. R. Luo, "Evaluation of a Model of Colour Vision by Magnitude Scalings: Discussion of Collected Results," *Color Research & Application*, vol. 19, no. 1, pp. 27–33, 1994.

[119] M. R. Luo, M.-C. Lo, and W.-G. Kuo, "The LLAB(*l*:*c*) Colour Model," *Color Research & Application*, vol. 21, no. 6, pp. 412–429, 1996.

[120] Commission Internationale de l'Éclairage, "A Colour-appearance Model for Colour Management Systems: CIECAM02," *CIE Technical Report 159:2004*, 2004.

[121] N. Moroney, M. D. Fairchild, R. G. W. Hunt, *et al.*, "The CIECAM02 Color-appearance Model," in *Proceedings of the 10th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2002, pp. 23–27.

[122] C. Li, M. R. Luo, R. G. W. Hunt, *et al.*, "The Performance of CIECAM02," in *Proceedings of the 10th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2002, pp. 28–32.

[123] "ICC Profiles, Color Appearance Modeling, and the Microsoft Windows Color System," in *Color Management: Understanding and using ICC Profiles*, P. Green and M. A. Kriss, Eds. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2010, pp. 53–55.

[124] I. Tastl, M. Bhachech, N. Moroney, *et al.*, "ICC Colour Management and CIECAM02," in *Proceedings of the 13th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2005, pp. 217–223.

[125] R. Gury and M. Shaw, "Dealing with Imaginary Color Encodings in CIECAM02 in an ICC workflow," in *Proceedings of the 13th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2005, pp. 318–318.

[126] R. W. G. Hunt, "Light and Dark Adaptation and Perception of Color," *Journal of the Optical Society of America*, vol. 42, no. 3, pp. 190–199, 1952.

[127] J. C. Stevens and S. S. Stevens, "Brightness Functions: Effects of Adaptation," *Journal of the Optical Society of America*, vol. 53, no. 3, pp. 375–385, 1963.

[128] C. J. Bartleson and E. J. Breneman, "Brightness Perception in Complex Fields," *Journal of the Optical Society of America*, vol. 57, no. 7, pp. 953–957, 1967.

[129] M. R. Luo, X. W. Gao, and S. A. R. Sciviner, "Quantifying Colour Appearance. Part V. Simultaneous Contrast," *Color Research & Application*, vol. 20, no. 1, pp. 18–28, 1995.

[130] M. R. Luo, G. Cui, and C. Li, "Uniform Colour Spaces Based on CIECAM02 Colour Appearance Model," *Color Research & Application*, vol. 31, no. 4, pp. 320–330, 2006.

[131] Wikipedia User Adoniscik, *Planckian locus in the CIE* 1960 *UCS using the Judd-Vos* 2° *CMF*, `https://en.wikipedia.org/wiki/CIE_1960_color_space`, [Online; accessed 16-April-2018], 2008.

[132] A. R. Robertson, "Computation of Correlated Color Temperature and Distribution Temperature," *Journal of the Optical Society of America*, vol. 58, no. 11, pp. 1528–1535, 1968.

[133] J. Schanda and M. Dányi, "Correlated Color Temperature Calculations in the CIE 1976 Chromaticity Diagram," *Color Research & Application*, vol. 2, no. 4, pp. 161–163, 1977.

[134] J. Schanda, M. Mészáros, and G. Czibula, "Calculating Correlated Color Temperature with a Desktop Programmable Calculator," *Color Research & Application*, vol. 3, no. 2, pp. 65–69, 1978.

[135] M. Krystek, "An Algorithm to Calculate Correlated Colour Temperature," *Color Research & Application*, vol. 10, no. 1, pp. 38–40, 1985.

[136] J. L. Gardner, "Correlated Colour Temperature Uncertainty and Estimation," *Metrologia*, vol. 37, no. 5, pp. 381–384, 2000.

[137] X. Qiu, "Formulas for Computing Correlated Color Temperature," *Color Research & Application*, vol. 12, no. 5, pp. 285–287, 1987.

[138] C. S. McCamy, "Correlated Color Temperature as an Explicit Function of Chromaticity Coordinates," *Color Research & Application*, vol. 17, no. 2, pp. 142–144, 1992.

[139] C. S. McCamy, "Correlated Color Temperature as an Explicit Function of Chromaticity Coordinates (Erratum)," *Color Research & Application*, vol. 18, no. 2, p. 150, 1993.

[140] J. Hernández-Andrés, R. L. Lee, and J. Romero, "Calculating Correlated Color Temperatures Across the Entire Gamut of Daylight and Skylight Chromaticities," *Applied Optics*, vol. 38, no. 27, pp. 5703–5709, 1999.

[141] C. Li, G. Cui, M. Melgosa, *et al.*, "Accurate Method for Computing Correlated Color Temperature," *Optics Express*, vol. 24, no. 13, pp. 14 066–14 078, 2016.

[142] B. T. Polyak, "Newton's Method and its Use in Optimization," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1086–1096, 2007.

[143] D. Kaye, *BrainyQuote.com: Quotes of Danny Kaye*, `https://www.brainyquote.com/quotes/danny_kaye_125475`, [Online; accessed 22-May-2018].

[144] J. P. Freyssinier and M. Rea, "A Two-metric Proposal to Specify the Color-rendering Properties of Light Sources for Retail Lighting," in *Proceedings of the SPIE 7784, 10th International Conference on Solid State Lighting*, San Diego, CA, USA: International Society for Optics and Photonics, 2010, pp. 77840V–1–77840V–6.

[145] M. Rea, L. Deng, and R. Wolsey, "Full-Spectrum Light Sources," *Lighting Answers*, vol. 7, no. 5, pp. 1–18, 2005.

[146] K. Hashimoto, T. Yano, M. Shimizu, *et al.*, "New Method of Specifying Color Rendering Properties of Light Sources Based on Feeling of Contrast," *Color Research & Application*, vol. 32, no. 5, pp. 361–371, 2007.

[147] W. A. Thornton, "Color-discrimination Index," *Journal of the Optical Society of America*, vol. 62, no. 2, pp. 191–194, 1972.

[148] S. A. Fotios, "The Perception of Light Sources of Different Colour Properties," PhD thesis, University of Manchester Institute of Science and Technology, 1997.

[149] M. R. Luo, "The Quality of Light Sources," *Coloration Technology*, vol. 127, no. 2, pp. 75–87, 2011.

[150] K. A. G. Smet, J. Schanda, and L. Whitehead, "CRI2012: A Proposal for Updating the CIE Colour Rendering Index," *Lighting Research & Technology*, vol. 45, no. 6, pp. 689–709, 2013.

[151] A. David, P. T. Fini, K. W. Houser, *et al.*, "Development of the IES Method for Evaluating the Color Rendition of Light Sources," *Optics Express*, vol. 23, no. 12, pp. 15 888–15 906, 2015.

[152] Commission Internationale de l'Éclairage, "Colour Fidelity Index for Accurate Scientific Use," *CIE Technical Report 224*, 2017.

[153] K. Smet, W. R. Ryckaert, M. R. Pointer, *et al.*, "Correlation Between Colour Quality Metric Predictions and Visual Appreciation of Light Sources," *Optics Express*, vol. 19, no. 9, pp. 8151–8166, 2011.

[154] S. Jost-Boissard, P. Avouac, and M. Fontoynont, "Assessing the Colour Quality of LED Sources: Naturalness, Attractiveness, Colourfulness and Colour Difference," *Lighting Research & Technology*, vol. 47, no. 7, pp. 769–794, 2015.

[155] P. Bodrogi, S. Brúckner, and T. Q. Khanh, "Ordinal Scale Based Description of Colour Rendering," *Color Research & Application*, vol. 36, no. 4, pp. 272–285, 2011.

[156] L. A. Whitehead and M. A. Mossman, "A Monte Carlo Method for Assessing Color Rendering Quality With Possible Application to Color Rendering Standards," *Color Research & Application*, vol. 37, no. 1, pp. 13–22, 2012.

[157] P. van der Burgt and J. van Kemenade, "About Color Rendition of Light Sources: The Balance Between Simplicity and Accuracy," *Color Research & Application*, vol. 35, no. 2, pp. 85–93, 2010.

[158] C. I. de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.2, 1974.

[159] Commission Internationale de l'Éclairage, *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, CIE Technical Report 13.3, 1995.

[160] C. Li, M. R. Luo, and C. Li, *Assessing Colour Rendering Properties of Daylight Sources Part II: A New Colour Rendering Index: CRI-CAM02UCS*, `https://www.researchgate.net/publication/268364092_Assessing_Colour_Rendering_Properties_of_Daylight_Sources_Part_II_A_New_Colour_Rendering_Index_CRI-CAM02UCS`, [Online; accessed 16-April-2018], 2009.

[161] H. Yaguchi, Y. Takahashi, and Y. Mizokami, "Categorical Colour Rendering Index Based on the CIECAM02," in *Proceedings of the 12th AIC Colour Congress*, Newcastle Gateshead, United Kingdom: Association Internationale de la Couleur (AIC), 2013, pp. 1441–1444.

[162] G. Wyszecki, "Proposal for a New Color-Difference Formula," *Journal of the Optical Society of America*, vol. 53, no. 11, pp. 1318–1319, 1963.

[163] Commission Internationale de l'Éclairage, *CIE Position Statement on CRI and Colour Quality Metrics*, `http://www.cie.co.at/sites/default/files/CIE%20Position%20Statement%20on%20CRI%20and%20Colour%20Quality%20Metrics%20V2.pdf`, [Online; accessed 16-April-2018], 2015.

[164] Commission Internationale de l'Éclairage, "Colour Rendering of White LED Light Sources," *CIE Technical Report 177 : 2007*, 2007.

[165] P. Bodrogi, P. Csuti, F. Szabó, *et al.*, "Why Does the CIE Colour Rendering Index Fail for White RGB LED Light Sources?" In *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 24–27.

[166] Y. Ohno, "Simulation Analysis of White LED Spectra and Color Rendering," in *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 28–32.

[167] Y. Ohno, "Spectral Design Considerations for Color Rendering of White LED Light Sources.," *Optical Engineering*, vol. 44, no. 11, pp. 111 302–111 310, 2005.

[168] W. Davis, J. L. Gardner, and Y. Ohno, "NIST Facility for Color Rendering Simulation," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 519–522.

[169] Y. Nakano, H. Tahara, K. Suehara, *et al.*, "Application of Multispectral Camera to Color Rendering Simulator," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 1625–1628.

[170] N. Sándor and J. Schanda, "CIE Visual Colour-rendering Experiments," in *Proceedings of the 10th AIC Colour Congress*, Granada, Spain: Association Internationale de la Couleur (AIC), 2005, pp. 511–514.

[171] N. Sándor, P. Csuti, P. Bodrogi, *et al.*, "Visual Observation of Colour Rendering," in *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*, Vienna, Austria: International Commission on Illumination CIE, 2004, pp. 16–19.

[172] N. Sándor, P. Bodrogi, P. Csuti, *et al.*, "Direct Visual Assessment of Colour Rendering," in *Proceedings of the 25th CIE Session*, Vienna, Austria: International Commission on Illumination CIE, 2003, pp. D1–42–D1–45.

[173] F. Szabó, N. Sándor, P. Bodrogi, *et al.*, "Colour rendering of white LED light sources: Visual experiment with colour samples simulated on a colour monitor," in *Proceedings of the CIE* 2005 *Midterm Meeting: Vision and Lighting in Mesopic Conditions*, Vienna, Austria: International Commission on Illumination CIE, 2005.

[174] D. B. Judd, "Hue Saturation and Lightness of Surface Colors with Chromatic Illumination," *Journal of the Optical Society of America*, vol. 30, no. 1, pp. 2–32, 1940.

[175] P. C. Mahalanobis, "On the Generalised Distance in Statistics," *Proceedings of the National Institute of Science in India*, vol. 2, no. 1, pp. 49–55, 1936.

[176] M. Sapp, F. Obiakor, A. J. Gregas, *et al.*, "Mahalanobis Distance: A Multivariate Measure of Effect in Hypnosis Research," *Sleep and Hypnosis*, vol. 9, no. 2, pp. 67–70, 2007.

[177] F. Ebner, "Derivation and Modelling Hue Uniformity and Development of the IPT Color Space," PhD thesis, Rochester Institute of Technology, 1998.

[178] F. Ebner and M. D. Fairchild, "Development and Testing of a Color Space (IPT) with Improved Hue Uniformity," in *Proceedings of the 6th Color and Imaging Conference: Color Science, Systems and Applications*, Scottsdale, AZ, USA: Society for Imaging Science and Technology (IS&T), 1998, pp. 8–13.

[179] K. W. Houser, M. Wei, A. David, *et al.*, "Review of Measures for Light-source Color Rendition and Considerations for a Two-measure System for Characterizing Color Rendition," *Optics Express*, vol. 21, no. 8, pp. 10 393–10 411, 2013.

[180] J. A. C. Yule, *Principles of Color Reproduction: Applied to Photomechanical Reproduction, Color Photography, and the Ink, Paper, and Other Related Industries*. New York, NY, USA: John Wiley & Sons Inc., 1967.

[181] K. Duncker, "The Influence of Past Experience upon Perceptional Properties," *American Journal of Psychology*, vol. 52, no. 2, pp. 255–265, 1939.

[182] J. S. Bruner, L. Postman, and J. Rodrigues, "Expectation and the Perception of Color," *American Journal of Psychology*, vol. 64, no. 2, pp. 216–227, 1951.

[183] R. S. Harper, "The Perceptual Modification of Colored Figures," *American Journal of Psychology*, vol. 66, no. 1, pp. 86–89, 1953.

[184] C. Witzel, H. Valkova, T. Hansen, *et al.*, "Object Knowledge Modulates Colour Appearance," *i-Perception*, vol. 2, no. 1, pp. 13–49, 2011.

[185] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, "Color Appearance of Familiar Objects: Effetcs of Shape, Texture, and Illumination Changes," *Journal of Vision*, vol. 8, no. 5, pp. 1–16, 2008.

[186] Y. Ling and A. C. Hurlbert, "Color and Size Interactions in a Real 3D Object Similarity Task," *Journal of Vision*, vol. 4, no. 9, pp. 721–734, 2004.

[187] R. S. Berns and M. E. Gorzynski, "Simulating Surface Colors on CRT Displays: The Importance of Cognitive Clues," in *Proceedings of the AIC Midterm Meeting: Colour and Light*, Sydney, Australia: Association Internationale de la Couleur (AIC), 1991, pp. 21–24.

[188] D. H. Brainard, "Color Constancy in the Nearly Natural Image Part 2: Achromatic Loci," *Journal of the Optical Society of America A*, vol. 15, no. 2, pp. 307–325, 1998.

[189] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 5–55, 1932.

[190] P. Bodrogi and T. Q. Khanh, *Illumination, Color and Imaging: Evaluation and Optimization of Visual Displays*, T. Lowe, Ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2012.

[191] Y. Zhu, M. R. Luo, S. Fischer, *et al.*, "Long-term Memory Color Investigation: Culture Effect and Experimental Setting Factors," *Journal of the Optical Society of America A*, vol. 34, no. 10, pp. 1757–1768, 2017.

[192] S. Fischer, P. Bodrogi, T. Q. Khanh, *et al.*, "Memory Colors Part I: Comparison of Chinese and German Observers," in *Proceedings of the LICHT Symposium 2016*, Karlsruhe: Karlsruher Institut für Technologie (KIT), 2016, pp. 102–110.

[193] Y. Zhu, M. Luo, L. Xu, *et al.*, "Investigation of Memory Colours Across Cultures," in *Proceedings of the 23rd Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2015, pp. 133–136.

[194] W. Davis and Y. Ohno, "Toward an Improved Color Rendering Metric," in *Proceedings of the SPIE 5941, 5th International Conference on Solid State Lighting*, Bellingham, WA, USA: International Society for Optics and Photonics, 2005, 59411G–1–59411G–8.

[195] S. Jost-Boissard, M. Fontoynont, and J. Blanc-Gonnet, "Perceived Lighting Quality of LED Sources for the Presentation of Fruit and Vegetables," *Journal of Modern Optics*, vol. 56, no. 13, pp. 1420–1432, 2009.

[196] M. Royer, A. Wilkerson, M. Wei, *et al.*, "Human Perceptions of Colour Rendition Vary with Average Fidelity, Average Gamut, and Gamut Shape," *Lighting Research & Technology*, vol. 49, no. 8, pp. 966–991, 2017.

[197] A. Werner, "Spatial and Temporal Aspects of Chromatic Adaptation and their Functional Significance for Colour Constancy," *Vision Research*, vol. 104, pp. 80–89, 2014.

[198] G. Derefeldt, T. Swartling, U. Berggrund, *et al.*, "Cognitive Color," *Color Research and Application*, vol. 29, no. 1, pp. 7–19, 2004.

[199] O. Rinner and K. R. Gegenfurtner, "Time Course of Chromatic Adaptation for Color Appearance and Discrimination," *Vision Research*, vol. 40, no. 14, pp. 1813–1826, 2000.

[200] S. Ishihara, *Ishihara's Tests for Colour Deficiency: 38 Plates Edition*. Tokyo: Kanehara Trading Inc., 2016.

[201] H. Ichikawa, K. Hukami, and S. Tanabe, *Standard Pseudoisochromatic Plates Part II: For Acquired Color Vision Defects*. Tokyo, New York: Igaku-Shoin, 1983.

[202] A. Linksz, "The Farnsworth Panel D-15 Test," *American Journal of Ophthalmology*, vol. 62, no. 1, pp. 27–37, 1966.

[203] Richmond Products Inc., *Farnsworth D-15 and Lanthony Test Instructions*, `http://www.richmondproducts.com/files/8113/1550/0538/FR_15_Farnsworth_and_LanthonyD15_Instructions_Rev_1.7_0506.pdf`, [Online; accessed 07-August-2017], 2006.

[204] S.-S. Guan and M. R. Luo, "Investigation of Parametric Effects Using Small Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 331–343, 1999.

[205] J. B. Kruskal, "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[206] A. P. M. Coxon, *The User's Guide to Multidimensional Scaling: With Special Reference to the MDS(X) Library of Computer Programs*. Portsmouth, NH, USA: Heinemann Educational Publishers, 1982.

[207] E. Coates, K. Y. Fong, and B. Rigg, "Uniform Lightness Scales," *Journal of the Society of Dyers and Colourists*, vol. 97, no. 4, pp. 179–183, 1981.

[208] S. Adler, K. P. Chaing, T. F. Chong, *et al.*, "Uniform Chromaticity Scales: New Experimental Data," *Journal of the Society of Dyers and Colourists*, vol. 98, no. 1, pp. 14–20, 1982.

[209] W. Schultze, "The Usefulness of Colour-Difference Formulae for Fixing Colour Tolerances," in *Proceedings of the Helmholtz Memorial Symposium on Color Metrics*, Soesterberg, Holland: Association Internationale de la Couleur (AIC), 1972, pp. 254–265.

[210] S.-S. Guan and M. R. Luo, "A Colour-Difference Formula for Assessing Large Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 344–355, 1999.

[211] S.-S. Guan and M. R. Luo, "Investigation of Parametric Effects Using Large Colour Differences," *Color Research and Application*, vol. 24, no. 5, pp. 356–368, 1999.

[212] H. Xu, H. Yaguchi, and S. Shioiri, "Estimation of Color-Difference Formulae at Color Discrimination Threshold Using CRT-generated Stimuli," *Optical Review*, vol. 8, no. 2, pp. 142–147, 2001.

[213] S. Shen and R. S. Berns, "Evaluating Color Difference Equation Performance Incorporating Visual Uncertainty," *Color Research and Application*, vol. 34, no. 5, pp. 375–390, 2009.

[214] R. S. Berns and B. X. Hou, "RIT-DuPont Supra-Threshold Color-tolerance Individual Color-difference Pair Dataset," *Color Research and Application*, vol. 35, no. 4, pp. 274–283, 2010.

[215] R. Shamey, L. M. Cárdenas, D. Hinks, *et al.*, "Comparison of Naïve and Expert Subjects on the Assessment of Small Color Differences," *Journal of the Optical Society of America A*, vol. 27, no. 6, pp. 1482–1489, 2010.

[216] J. Ma, H. S. Xu, M. R. Luo, *et al.*, "Color Appearance and Visual Measurements for Color Samples with Gloss Effect," *Chinese Optics Letters*, vol. 7, no. 9, pp. 869–872, 2009.

[217] S. G. Kandi and M. A. Tehrani, "Investigating the Effect of Texture on the Performance of Color Difference Formulae," *Color Research and Application*, vol. 35, no. 2, pp. 94–100, 2010.

[218] Z. N. Huang, H. S. Xu, M. R. Luo, *et al.*, "Assessing Total Differences for Effective Samples Having Variations in Color, Coarseness, and Glint," *Chinese Optics Letters*, vol. 8, no. 7, pp. 717–720, 2010.

[219] H. Wang, G. Cui, M. R. Luo, *et al.*, "Evaluation of Colour-Difference Formulae for Different Colour-Difference Magnitudes," *Color Research & Application*, vol. 37, no. 5, pp. 316–325, 2012.

[220] M. Melgosa, P. A. García, L. Gómez-Robledo, *et al.*, "Notes on the Application of the Standardized Residual Sum of Squares Index for the Assessment of Intra- and Inter-observer Variability in Color-Difference Experiments," *Journal of the Optical Society of America A*, vol. 28, no. 5, pp. 949–953, 2011.

[221] P. E. Shrout and J. L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.

[222] R. W. Pridmore and M. Melgosa, "Effect of Luminance of Samples on Color Discrimination Ellipses: Analysis and Prediction of Data," *Color Research and Application*, vol. 30, no. 3, pp. 186–197, 2005.

[223] W. R. J. Brown, "The Influence of Luminance Level on Visual Sensitivity to Color Differences," *Journal of the Optical Society of America*, vol. 41, no. 10, pp. 684–688, 1951.

[224] H. Xu, H. Yaguchi, and S. Shioiri, "Chromaticity-Discrimination Ellipses for Surface Colours," *Color Research and Application*, vol. 11, no. 1, pp. 25–42, 1986.

[225] M. Melgosa, M. M. Pérez, and A. El Moraghi, "Color Discrimination Results from a CRT Device: Influence of Luminance," *Color Research and Application*, vol. 24, no. 1, pp. 677–679, 1999.

[226] F. Carreño and J. M. Zoido, "The Influence of Luminance on Color-Difference Thresholds," *Color Research and Application*, vol. 26, no. 5, pp. 362–368, 2001.

[227] R. S. Berns, D. H. Alman, L. Reniff, *et al.*, "Visual Determination of Suprathreshold Color-Difference Tolerances Using Probit Analysis," *Color Research and Application*, vol. 16, no. 5, pp. 297–316, 1991.

[228] M. Melgosa, E. Hita, A. J. Poza, *et al.*, "Suprathreshold Color-Difference Ellipsoids for Surface Colors," *Color Research and Application*, vol. 22, no. 3, pp. 148–155, 1997.

[229] A. Yebra, J. A. García, J. L. Nieves, *et al.*, "Chromatic Discrimination in Relation to Luminance Level," *Color Research and Application*, vol. 26, no. 2, pp. 123–131, 2001.

[230] H. Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, vol. 2, no. 3, pp. 360–378, 1931.

[231] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Harlow, United Kingdom: Pearson Education Limited, 2014.

[232] G. E. P. Box, "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, vol. 36, no. 3/4, pp. 317–346, 1949.

[233] H. C. Box, *Set Lighting Technician's Handbook: Film Lighting Equipment, Practice, and Electrical Distribution*. Burlington, MA, USA: Taylor & Francis, 2010.

[234] Commission Internationale de l'Éclairage, "Colorimetry – Part 6: CIEDE2000 Colour-difference Formula," *ISO/CIE 11664-6:2014*, 2014.

[235] H. W. Lilliefors, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[236] B. Rosner, "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, 1983.

[237] B. Iglewicz and D. C. Hoaglin, *How to Detect and Handle Outliers*. Milwaukee, WI: ASQC Quality Press, 1993.

[238] F. E. Grubbs, "Sample Criteria for Testing Outlying Observations," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 27–58, 1950.

[239] F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[240] A. N. Kolmogorov, "Sulla Determinazione Empirica di una Legge di Distribuzione," *Giornale dell'Istituto Italiano degli Attuari*, vol. 4, pp. 83–91, 1933.

[241] D. Öztuna, A. H. Elhan, and E. Tüccar, "Investigation of Four Different Normality Tests in Terms of Type 1 Error rate and Power under Different Distributions," *Turkish Journal of Medical Sciences*, vol. 36, no. 3, pp. 171–176, 2006.

[242] W. D Wayne, "Kolmogorov–Smirnov One-Sample Test," in *Applied Nonparametric Statistics*. Boston, MA, USA: PWS-Kent, 1990, pp. 319–330.

[243] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[244] S. S. Shapiro and R. S. Francia, "An Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215–216, 1972.

[245] N. M. Razali and B. W. Yap, "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[246] A. Ghasemi and S. Zahediasl, "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012.

[247] H.-Y. Kim, "Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis," *Restorative Dentistry and Endodontics*, vol. 38, no. 1, pp. 52–54, 2013.

[248] D. N. Joanes and C. A. Gill, "Comparing Measures of Sample Skewness and Kurtosis," *The Statistician*, vol. 47, no. 1, pp. 183–189, 1998.

[249] D. Cramer, *Basic Statistics for Social Research: Step-by-step Calculations and Computer Techniques Using Minitab*. Abingdon-on-Thames, United Kingdom: Routledge Taylor & Francis Group, 1997.

[250] J. F. Ratcliffe, "The Effect on the *t* Distribution of Non-normality in the Sampled Population," *Applied Statistics*, vol. 17, no. 1, pp. 42–48, 1968.

[251] S. S. Sawilowsky and R. C. Blair, "A More Realistic Look at the Robustness and Type II Error Properties of the *t* Test to Departures from Population Normality," *Psychological Bulletin*, vol. 111, no. 2, pp. 352–360, 1992.

[252] S. S. Sawilowsky and S. B. Hillman, "Power of the Independent Samples *t* Test under a Prevalent Psychometric Measure Distribution," *Journal of Consulting and Clinical Psychology*, vol. 60, no. 2, pp. 240–243, 1992.

[253] H. O. Posten, "Robustness of the Two-Sample *t* Test," in *Robustness of Statistical Methods and Nonparametric Statistics*, D. Rasch and M. L. Tiku, Eds. Dordrecht, Holland: D. Reidel Publishing Company, 1984, pp. 92–99.

[254] W. S. Gosset (Student), "The Probable Error of a Mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.

[255] E. Kreyszig, *Introductory Mathematical Statistics: Principles and Methods*. New York: John Wiley & Sons, 1970.

[256] D. G. Altman, *Practical Statistics for Medical Research*. Boca Raton, FL, USA: Chapman & Hall, 1991.

[257] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[258] H. B. Mann and D. R. Whitney, "On a Test of Whether one of two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[259] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd ed. New York: John Wiley & Sons, 2014.

[260] A. J. Vickers, "Parametric Versus Non-parametric Statistics in the Analysis of Randomized Trials with Non-normally Distributed Data," *BMC Medical Research Methodology*, vol. 5, no. 1, pp. 35–47, 2005.

[261] M. B. Brown and A. B. Forsythe, "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 364–367, 1974.

[262] H. Levene, "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, S. G. Ghurye, W. Hoeffding, *et al.*, Eds. Stanford, CA, USA: Stanford University Press, 1960, pp. 278–292.

[263] K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*. New York: John Wiley & Sons, 1965.

[264] W. U. Behrens, "Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen," *Landwirtschaftliche Jahrbücher*, vol. 68, pp. 807–837, 1929.

[265] R. A. Fisher, "The Fiducial Argument in Statistical Inference," *Annals of Human Genetics*, vol. 6, no. 4, pp. 391–398, 1935.

[266] R. A. Fisher, "The Asymptotic Approach to Behrens' Integral with further Tables for the *d* Test of Significance," *Annals of Human Genetics*, vol. 11, no. 1, pp. 141–172, 1941.

[267] T. P. Hettmansperger and J. W. McKean, *Robust Nonparametric Statistical Methods*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2011.

[268] B. L. Welch, "The Significance of the Difference Between Two Means when the Population Variances are Unequal," *Biometrika*, vol. 29, no. 3/4, pp. 350–362, 1938.

[269] B. L. Welch, "The Generalization of "student's"problem when Several Different Population Variances are Involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.

[270] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. San Diego, CA, USA: GraphPad Software, Inc, 2003.

[271] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, *et al.*, "Modeling of the Bacterial Growth Curve," *Applied and Environmental Microbiology*, vol. 56, no. 6, pp. 1875–1881, 1990.

[272] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

[273] R. M. Warner, *Applied Statistics: From Bivariate Through Multivariate Techniques*, 2nd. London, United Kingdom: SAGE Publications, Ltd, 2013.

[274] J. E. Pustejovsky, "Converting from d to r to z when the Design uses Extreme Groups, Dichotomization, or Experimental Control," *Psychol. Methods*, vol. 19, no. 1, pp. 92–112, 2014.

[275] M. Borenstein, "Effect Sizes for Continuous Data," in *The Handbook of Research Synthesis and Meta-Analysis*, H. Cooper, L. V. Hedges, and J. C. Valentine, Eds., 2nd. New York, NY, USA: Sage Foundation, 2009, pp. 221–235.

[276] A. Field, *Discovering Statistics using IBM SPSS Statistics*, 4th. London, United Kingdom: SAGE Publications, Inc., 2013.

[277] J. Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis using IBM SPSS*, 6th. Philadelphia, PA, USA: Mcgraw-Hill Higher Education, Open University Press, 2016.

[278] R. R., "Parametric measures of effect size," in *The Handbook of Research Synthesis*, H. Cooper and L. V. Hedges, Eds. New York, NY, USA: Russel SAGE Foundation, 1994, pp. 231–244.

[279] J. P. Stevens, *Applied Multivariate Statistics for the Social Science*, 5th. New York, NY, USA: Taylor & Francis Group, 2009.

[280] I. T Hernádvölgyi, *Generating Random Numbers from the Multivariate Normal Distribution*, `ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia013_2s09/material_de_apoio/gen_rand_multivar.pdf`, [Online; accessed 14-March-2018], 1998.

[281] C. Spearman, "The Proof and Measurement of Association between two Things," *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[282] W. D Wayne, "Spearman Rank Correlation Coefficient," in *Applied Nonparametric Statistics*. Boston, MA, USA: PWS-Kent, 1990, pp. 358–365.

[283] A. Rohatgi, *WebPlotDigitizer 4.0*, `https://automeris.io/WebPlotDigitizer/`, [Online; accessed 28-March-2018], 2017.

[284] M. Wei, K. W. Houser, G. R. Allen, *et al.*, "Color Preference under LEDs with Diminished Yellow Emission," *Journal of the Illuminating Engineering Society*, vol. 10, no. 3, pp. 119–131, 2014.

[285] F. Szabó, R. Kéri, J. Schanda, *et al.*, "A Study of Preferred Colour Rendering of Light Sources: Home Lighting," *Lighting Research & Technology*, vol. 48, no. 2, pp. 103–125, 2016.

[286] J. VanRie, "The Effect of the Spectral Composition of a Light Source on the Visual Appreciation of a Composite Object Set," in *Technical Report to the User Committee of the IWT-TETRA Project (80163), Appendix* 4, Diepenbeek, Belgium, 2009.

[287] PennState, Eberly College of Science, *STAT 503 Design of Experiments: The Latin Square Design*, `https://onlinecourses.science.psu.edu/stat503/node/21`, [Online; accessed 27-March-2018].

[288] Y. Imai, T. Kotani, and T. Fuchida, "A Study of Colour Rendering Properties based on Colour Preference in Adaptation to LED Lighting," in *Proceedings of the CIE 2012 Conference: Lighting Quality and Energy Efficiency*, Vienna, Austria: International Commission on Illumination CIE, 2012, pp. 369–374.

[289] Y. Imai, T. Kotani, and T. Fuchida, "A Study of Colour Rendering Properties based on Colour Preference of Objects in Adaptation to LED Lighting," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 62–67.

[290] S. Jost and M. Fontoynont, "Colour Rendering of Face Complexion and Hair under LED Sources," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 53–61.

[291] A. Tsukitani, "Optimization of Colour Quality for Landscape Lighting based on Feeling of Contrast Index," in *CIE Centenary Conference: Towards a new Century of Light*, Vienna, Austria: International Commission on Illumination CIE, 2013, pp. 68–71.

[292] Y. Lin, J. He, A. Tsukitani, *et al.*, "Colour Quality Evaluation of Natural Objects based on the Feeling of Contrast Index," *Lighting Research & Technology*, vol. 48, no. 3, pp. 323–339, 2016.

[293] K. W. Houser, D. K. Tiller, and X. Hu, "Tuning the Fluorescent Spectrum for the Trichromatic Visual Response: A Pilot Study," *Journal of the Illuminating Engineering Society*, vol. 1, no. 1, pp. 7–23, 2005.

[294] Z. Huang, Q. Liu, S. Westland, *et al.*, "Light Dominates Colour Preference when Correlated Colour Temperature Differs," *Lighting Research & Technology*, 2017, [online-first].

[295] N. Narendran and L. Deng, "Color Rendering Properties of LED Light Sources," *Proceedings of SPIE*, vol. 4776, pp. 61–67, 2002.

[296] Q. Wang, H. Xu, F. Zhang, *et al.*, "Influence of Color Temperature on Comfort and Preference for LED Indoor Lighting," *Optik*, vol. 129, pp. 21–29, 2017.

[297] Q. Liu, Z. Huang, K. Xiao, *et al.*, "Gamut Volume Index: A Color Preference Metric based on Meta-analysis and Optimized Color Samples," *Optics Express*, vol. 25, no. 14, pp. 16 378–16 391, 2017.

[298] J. E. Hunter and F. L. Schmidt, *Methods of Meta-analysis: Correcting Error and Bias in Research findings*, 2nd. Thousand Oaks, CA, USA: SAGE Publications, 2004.

[299] J. Sánchez-Meca and F. Marín-Martínez, "Confidence Intervals for the Overall Effect Size in Random-Effects Meta-Analysis," *Psychological Methods*, vol. 13, no. 1, pp. 31–48, 2008.

[300] K. Sidik and J. N. Jonkman, "A Comparison of Heterogeneity Variance Estimators in Combining Results of Studies," *Statistics in Medicine*, vol. 26, no. 9, pp. 1964–1981, 2007.

[301] W. Viechtbauer, "Bias and Efficiency of Meta-analytic Variance Estimators in the Random-Effects Model," *Journal of Educational and Behavioral Statistics*, vol. 30, no. 3, pp. 261–293, 2005.

[302] G. Y. Zou, "Toward Using Confidence Intervals to Compare Correlations," *Psychological Methods*, vol. 12, no. 4, pp. 399–413, 2007.

[303] R. A. Fisher, "Frequency Distribution of the Values of the Correlation Coefficient in Samples of an Indefinitely Large Population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[304] R. A. Fisher, "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," *Metron*, vol. 1, pp. 3–32, 1921.

[305] Z. Ugray, L. Lasdon, J. Plummer, *et al.*, "Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization," *INFORMS Journal on Computing*, vol. 19, no. 3, pp. 328–340, 2007.

[306] P. Csuti and J. Schanda, "Colour Matching Experiments with RGB-LEDs," *Color Research & Application*, vol. 33, no. 2, pp. 108–112, 2008.

[307] P. Csuti and J. Schanda, "A Better Description of Metameric Experience of LED Clusters," *Light & Engineering*, vol. 18, no. 1, pp. 44–50, 2010.

[308] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 1: Experiments in a Room," *Lighting Research & Technology*, vol. 49, no. 6, pp. 697–713, 2017.

[309] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 2: Experiments in a Viewing Booth and Analysis of the Combined Dataset," *Lighting Research & Technology*, vol. 49, no. 6, pp. 714–726, 2017.

[310] T. Q. Khanh and P. Bodrogi, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 3: Experiments with Makeup Products and Analysis of the Complete Warm White Dataset," *Lighting Research & Technology*, vol. 50, no. 2, pp. 218–236, 2018.

[311] T. Q. Khanh, P. Bodrogi, Q. T. Vinh, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 4: Experiments with Still Life Arrangements at Different Correlated Colour Temperatures," *Lighting Research & Technology*, 2017, [online-first].

[312] T. Q. Khanh, P. Bodrogi, X. Guo, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 5: A Colour Preference Experiment at 2000 lx in a Real Room," *Lighting Research & Technology*, 2017, [online-first].

[313] E. Schrödinger, *BrainyQuote.com: Quotes of Erwin Schrödinger*, `https://www.brainyquote.com/quotes/erwin_schrodinger_304801`, [Online; accessed 30-May-2018].

OWN PUBLICATIONS

Some ideas and figures presented in this thesis may have appeared previously in one of the following publications:

Journal Publications:

[1] S. Babilon and T. Q. Khanh, "Color Appearance Rating of Familiar Real Objects Under Immersive Viewing Conditions," *Color Research & Application*, vol. 43, no. 4, pp. 551–568, 2018.

[2] Y. Zhu, M. R. Luo, S. Babilon, *et al.*, "Long-term Memory Color Investigation: Culture Effect and Experimental Setting Factors," *Journal of the Optical Society of America A*, vol. 34, no. 10, pp. 1757–1768, 2017.

[3] T. Q. Khanh, P. Bodrogi, S. Babilon, *et al.*, "Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 5: A Colour Preference Experiment at 2000 lx in a Real Room," *Lighting Research & Technology*, 2017, [published as online-first].

[4] P. Bodrogi, X. Guo, S. Babilon, *et al.*, "Observer Preference for Perceived Illumination Chromaticity," *Color Research & Application*, vol. 43, no. 4, pp. 506–516, 2018.

[5] S. Babilon and T. Q. Khanh, "Memory Colors and the Assessment of Color Quality in Lighting Applications," *Journal of the Illuminating Engineering Society*, 2018, [under peer review].

[6] S. Babilon and T. Q. Khanh, "Impact of the Adapted White Point and the Cultural Background on Memory Color Assessments," *Optics Express*, 2018, [under peer review].

[7] S. Babilon, P. Myland, J. Klabes, *et al.*, "Spectral Reflectance Estimation of Organic Tissue for Improved Color Correction of Video Assisted Surgery," *Journal of Electronic Imaging*, 2018, [accepted, will be published soon].

[8] H. D. Potter, S. Babilon, P. Maris, *et al.*, "*Ab Initio* Study of Neutron Drops with Chiral Hamiltonians," *Physics Letters B*, vol. 739, pp. 445–450, 2014.

[9] S. Babilon and W. Marques Jr., "Sound Absorption and Dispersion in Dilute Polyatomic Gases: A Generalized Kinetic Approach," *Journal of Statistical Mechanics: Theory and Experiment*, P08004, 2013.

Peer-Reviewed Conference Contributions:

[1] S. Babilon and T. Q. Khanh, "Color Reproduction of Digital Camera Systems Using LED Spotlight Illumination," in *Proceedings of the 23rd Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2015, pp. 143–147.

[2] S. Babilon, P. Myland, M. Szarafanowicz, *et al.*, "Strengths and Limitations of a Uniform 3D-LUT Approach for Digital Camera Characterization," in *Proceedings of the 24th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2016, pp. 315–322.

[3] Y. Zhu, M. R. Luo, S. Babilon, *et al.*, "Investigation of Memory Colours Across Cultures," in *Proceedings of the 23rd Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2015, pp. 133–136.

[4] Y. Zhu, M. R. Luo, S. Babilon, *et al.*, "The Effectiveness of Colour Appearance Attributes for Enhancing Image Preference and Naturalness," in *Proceedings of the 24th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, Springfield, VA, USA: Society for Imaging Science and Technology (IS&T), 2016, pp. 231–236.

Non-Peer-Reviewed Conference Contributions:

[1] S. Babilon, Y. Jiang, T. Q. Khanh *et al.*, "Display-based Methods for Inter-Cultural Image Quality Evaluation," in *Proceedings of the CIE 2017 Midterm Meeting: Smarter Lighting for Better Life*, Vienna, Austria: Commission Internationale de l'Éclairage (CIE), 2017, pp. 746–757.

[2] S. Babilon, P. Bodrogi, T. Q. Khanh, *et al.*, "Memory Colors Part I: Comparison of Chinese and German Observers," in *Proceedings of the LICHT Symposium 2016*, Karlsruhe, Germany: Karlsruher Institut für Technologie (KIT), 2016, pp. 102–110.

[3] S. Babilon, R. Roth, A. Calci, *et al.*, "*Ab Initio* Study of Neutron Drops with Chiral NN+3N Interactions," in *Proceedings of the DPG Frühjahrstagung 2014: Physik der Hadronen und Kerne*, Frankfurt, Germany: Deutsche Physikalische Gesellschaft (DPG), 2014.

[4] J. P. Vary, P. Maris, S. Babilon, *et al.*, "*Ab Initio* No Core Shell Model – Recent Results and Further Prospects," in *Proceedings of the International Conference on the Nuclear Theory in the Supercomputing Era 2014 (NTSE-2014)*, Khabarovsk, Russia: Pacific National University Khabarovsk, 2016, pp. 154–165.

Invited Talks:

[1] S. Babilon, "Vom Licht zum Bild – Detailanalyse des Kamera-Workflows im Digitalen Zeitalter," in *Clustersymposium User Interfaces – Technologien, Komponenten, Fertigung und Test*, Ingolstadt, Germany: Technische Hochschule Ingolstadt (THI), 2016.

[2] S. Babilon, "Farberinnerung – Literaturstand und Forschungsergebnisse 1950er bis heute," in *DfwG Jahrestagung 2017*, Aschheim, Germany: Deutsche farbwissenschaftliche Gesellschaft e.V. (DfwG), 2017.

[3] S. Babilon and T. Q. Khanh, "Improvements on Memory-Based Approaches to Evaluate the Colour Rendition of White Light Sources," in *16th International Symposium on the Science and Technology of Lighting*, Sheffield, UK: University of Sheffield, 2018.

[4] S. Babilon and T. Q. Khanh, "Color Perception and Color Quality of White LED Light Sources," in *The Pigment and Color Science Forum 2018*, Boston, MA, USA: Smithers Rapra, 2018.

Supervised Theses:

[1] X. Xing, "*Estimation of Spectral Reflectance Using a Digital Camera System*," Master-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2016.

[2] T. M. Tanksale, "*Development of a Spectral Acquisition System Using a Tunable Light Source and a Trichromatic Camera*," Master-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2015.

[3] D. Stojanovic, "*Wahrnehmung und Präferenz von Weißtönen für die LED Innenraumbeleuchtung*," Master-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2017.

[4] Y. Jiang, "*Displaybasierte Methoden zur Untersuchung der Farbpräferenz von Chinesischen und Deutschen Probanden*," Master-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2017.

[5] C. Becker, "*Rekonstruktion von Tageslichtspektren mit Hilfe eines Farbsensors*," Bachelor-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2017.

[6] D. Sylla, "*Weiterentwicklung eines Monochromatoraufbaus zur Spektralen Sensorcharakterisierung*," Bachelor-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2016.

[7] P. J. Myland, "*Kolorimetrische Farbreproduktion Digitaler Kamerasysteme unter Verwendung Linearer Interpolationsmethoden*," Bachelor-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2016, [Gewinn des DfwG Förderpreises 2016].

[8] J. M. Klabes, "*Anwendung und Vergleich unterschiedlicher Charakterisierungsalgorithmen für hochqualitative Flüssigkristallbildschirme*," Bachelor-Thesis, Darmstadt, Germany: Technische Universität Darmstadt, 2016.

CURRICULUM VITAE

Persönliche Daten

Sebastian Babilon, geb. Fischer
Robert-Schneider-Str. 31
64289 Darmstadt

Tel.: (+49) 151 / 40 32 81 22
E-Mail: babilon@lichttechnik.tu-darmstadt.de

Geb. am 26.11.1987 in Lichtenfels
verheiratet, deutsch

Beruf

09/2014–heute **TU Darmstadt, Fachgebiet Lichttechnik**
Hochschulstr. 4a, 64289 Darmstadt

Wissenschaftlicher Mitarbeiter

- Leitung diverser Forschungsprojekte zu den Themen Farb- und Lichtqualität, Kamertechnologie sowie spektraler Rekonstruktion
- Betreuung von Lehrveranstaltungen und studentischen Arbeiten
- Organsiation von Tagungen und Seminaren der Lichttechnik sowie fachgebietsinterne Organisation
- Förder- und Drittmitteleinwerbung
- Lichttechnische Dienstleistungen sowie Durchführung von Auftragsmessungen und Begutachtungen

Bildung

Schule

09/1998–07/2007 **Abitur**
Arnold-Gymnasium Neustadt b. Coburg

Studium

10/2008–09/2011 **B.Sc. Physik**
Technische Universität Darmstadt

10/2011–03/2014 **M.Sc. Physik, Schwerpunkt Kern- und nukleare Astrophysik**
Technische Universität Darmstadt

Auslandsstudium

07/2012–02/2013 **Studienarbeit zur Theorie ein- und mehratomiger Gase**
Departamento de Física, Universidade Federal do Paraná, Brasilien

01/2014–03/2014 **Forschungspraktikum: Properties of Trapped Neutrons Interacting with Realistic Nuclear Hamiltonians**
Iowa State University, Nuclear Theory Group, Ames, IA, USA

Sonstiges

10/2011–03/2014 **Stipendiat der Studienstiftung des Deutschen Volkes**

04/2014–08/2014 **Stipendiat der Helmholtz Graduate School for Hadron and Ion Research**

Forschungsarbeiten

Bachelor-Thesis *"Brillouin-Scattering in Water: The Temperature Dependence of Linewidth"*
Institut für Angewandte Physik, Laser und Quantenoptik

Studienarbeit *"Sound Propagation and Dispersion in Dilute Polyatomic Gases: A Generalized Kinetic Approach"*
Departamento de Física, Universidade Federal do Paraná, Brasilien

Master-Thesis *"Ab-initio Study of Neutron Drops with Chiral NN+3N Interactions"*
Institut für Kernphysik, Theoretical Nuclear Physics ++

Arbeitserfahrung

10/2007–09/2008 **ASB Coburg Kranken- und Behindertenfahrdienst**
Parkstr. 6, 96450 Coburg

Zivildienstleistender/Fahrer

- Dialyse-, Arztbesuchs- und Krankenhausfahrten
- Beförderung geistig und körperlich behinderter Mitmenschen zur Mobilitätssteigerung
- Fahrten für Inklusionskindergärten und -schulen

09/2009–06/2012 **ulogistics transport UG**
Im Weihereck 30, 64342 Seeheim-Jugenheim

Fahrer/Transportdienstleister

- Regionale Stückgutverteilung von Printmedien
- Beförderungs- und betriebssichere Verladung der Güter gemäß StVO und StVZO
- Verladen mit Hilfe eines Flurförderzeuges und den dazugehörigen Zurr- und Anschlagmitteln
- Aus- und Anlieferung der Güter im gesamten Rhein-Main-Gebiet

04/2009–09/2009 **TU Darmstadt, Institut für Angewandte Physik**
Laser und Quantenoptik
Schloßgartenstr. 7, 64289 Darmstadt

Übungsgruppenleiter Physik I

- Direkte Übungsbetreuung der Studierenden in den wöchentlichen stattfindenden Präsenzübungen zur Vertiefung des Vorlesungsstoffs
- Abhalten von Sprechstunden zur Klärung etwaiger Fragen bezüglich Vorlesung/Übungen/Hausaufgaben
- Hausaufgabenkorrektur
- Mitgestalten, Aufsicht und Korrektur der Semestralklausuren

10/2009–03/2010 **TU Darmstadt, Institut für Kernphysik**
Laser- und Plasmaphysik
Schloßgartenstr. 2, 64289 Darmstadt

Übungsgruppenleiter Physik für Chemiker

- siehe oben

04/2010–03/2011 **TU Darmstadt, Institut für Angewandte Physik**
Laser und Quantenoptik
Schloßgartenstr. 7, 64289 Darmstadt

Übungsgruppenleiter Physik II und III

- siehe oben

04/2011–09/2011 **TU Darmstadt, Institut für Kernphysik**
Experimentelle Kernphysik mit radioaktiven Strahlen
Schloßgartenstr. 2, 64289 Darmstadt

Übungsgruppenleiter Physik IV

- siehe oben

10/2011–03/2014 **TU Darmstadt, Institut für Angewandte Physik**
Theoretische Quantenphysik
Hochschulstr. 4a, 64289 Darmstadt

Übungsgruppenleiter Theoretische Physik II-IV

- siehe oben

04/2013–09/2013 **TU Darmstadt, Institut für Kernphysik**
Theoretical Nuclear Physics ++
Schloßgartenstr. 2, 64289 Darmstadt

Übungsgruppenleiter Höhere Quantenmechanik

- siehe oben

03/2013–09/2013 **TU Darmstadt, Hochschulrechenzentrum**
Mornewegstr. 30, 64293 Darmstadt

Studentische Hilfskraft im Bereich E-Learning

- Durchführung von Beratungsgesprächen und Schulungen zum Thema E-Learning
- Vor-Ort Betreuung der Lehrenden bei der Umsetzung von E-Learning in Vorlesungen und Seminaren
- Technischer Support während des Vorlesungsbetriebs

06/2014–03/2015 **GSI Helmholtzzentrum für**
Schwerionenforschung GmbH
Planckstr. 1, 64291 Darmstadt

Besucherführer

- Presse- und Öffentlichkeitsarbeit
- Betreuung von Besuchergruppen inklusive Führungen durch das Forschungszentrum

Sprachen

Muttersprache	**Deutsch**
Sonst. Sprachen	**Englisch** – Lesen: fließend, Schreiben: verhandlungssicher, Sprechen: verhandlungssicher
	Portugiesisch – Lesen: erweiterte Grundkenntnisse, Schreiben: Grundkenntnisse, Sprechen: erweiterte Grundkenntnisse
	Französisch – Lesen: erweiterte Grundkenntnisse, Schreiben: Grundkenntnisse, Sprechen: Grundkenntnisse
Sprachreisen	**Southampton, England, 01.09.-21.09.2013** Southampton English Language Centre

Publikationen

Übersicht siehe Seite 221

Darmstadt, 26. Juni 2018

gez. Sebastian Michael Babilon

DECLARATION

---

§ 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Disserationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 26. Juni 2018

gez. Sebastian Michael Babilon

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Final Version as of June 26, 2018 (`classicthesis` version 4.2).