The influence of the skin colour on the perceived attributes

Mengmeng Wang

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Design

September 2017

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The publication related to the Chapter 3 experimental and Chapter 4 skin colour database:

Mengmeng WANG, Kaida XIAO, Ming Ronnier LUO, Yuteng ZHU, Sophie WUERGER, (2015) INVESTIGATION OF UNCERTAINTY OF SKIN COLOUR MEASUREMENTS. CIE conference proceeding 2015 Part 1. PP.294-299

Mengmeng WANG, Kaida XIAO, Vien CHEUNG, Sophie WUERGER, Ming Ronnier LUO (2015). COLOUR MEASUREMENT ON HUMAN SKIN. CIC conference proceeding 2015. pp. 230-234(5)

Mengmeng WANG, Kaida XIAO, Sophie WUERGER, Haoxue LIU, Ming HUANG, Ming Ronnier LUO (2016) SKIN COLOUR MEASUREMENT BY USING NON-CONTACT METHODS. CIE conference proceeding 2016. PP. 487-492

Mengmeng WANG, Ming Ronnier LUO, Kaida XIAO, Sophie WUERGER, Yuzhao WANG, Minchen WEI (2016) TWO NEW SPECTRAL DATABASES FOR SKIN COLOURS. CIC conference proceeding 2016. ISBN: 978-0-89208-325-1. PP. 266-270

Kaida XIAO, Mengmeng WANG, Jingjing YIN, Changjun LI, Sophie WUERGER (2015) DEVELOPMENT OF A SKIN REFLECTANCE RE-CONSTRUCTION MODEL USING A SKIN DATABASE. AIC conference proceeding 2015.

Yuzhao WANG, Ming Ronnier LUO, Mengmeng WANG, Kaida XIAO, Michael POINTER (2017) SPECTROPHOTOMETRIC MEASUREMENT OF HUMAN SKIN COLOUR. Color Research and Application (2017).

The publication related to the Chapter 5 facial skin colour and impressions:

Mengmeng WANG, Ming Ronnier LUO, Kaida XIAO (2017) THE IMPACT OF SKIN COLOUR ON FACIAL IMPRESSIONS. AIC Conference 2017. [Accepted May 2017]

Acknowledgement

The present study would never have been possible without the supervision from Professor Ming Ronnier Luo and Dr Vien Cheung, and advice from Dr Kaida Xiao, Professor Sophie Wuerger (University of Liverpool) and Professor Stephen Westland. I am so grateful to work with you in the past four years. Without your guidance and generous help, the research can never be accomplished. With a special mention to Dr Michael Pointer, he gave me the great support and endless help for my thesis. This thesis can never accomplish without your and Ronnier's guide.

Special acknowledgements to the International student scholarship of University of Leeds and Colour Group (GB) who provided financial support to my research and to funded me to participate conference. Without your support, this research cannot be carried out.

I want to bring my special thanks to all the research students and staffs in School of Design, University of Leeds; Psychology Department, University of Liverpool; Colour engineer Lab, Zhejiang University, whom give me endless support and help to my research. It was a great pleasure and honour to work with you.

I'm grateful to all of those who had voluntary participated the experiments at the University of Liverpool, Zhejiang University and Hangzhou Dianzi University. Their support helped this research to be accomplished. I am so grateful for all your understanding and support

The last never means the least, I would like to bring my thanks to my family and friends who encouraged and supported me in the past four years.

Abstract

Skin colour data are important for many applications such as medical, imaging, cosmetics. The present study was aimed to collect a comprehensive skin colour database, and to study the impact of the skin colour on the variety of facial impression attributes. Although many researchers and engineers have collected skin data，few of them studied the skin colours to measure the same locations on a large number of subjects from different ethnic groups using the same colour measuring instruments. As for studying the impact of the skin colour on the visual perceptions, many studies investigated the impact of the skin colour on the attractiveness, health and youth. Limited previous studies investigated the impact of the skin colour on the other impression attributes.

The present study was divided into two experiments, Experiments 1 and 2. Experiment 1 was to accumulate the skin colour database, named the Leeds Liverpool skin colour (LLSC). It included skin colours of 188 people from four ethnic groups (Caucasian, Oriental, South Asian and African) and both genders. Three colour measuring methods were used to accumulate the skin colour of each subject's 10 locations including facial locations (forehead, cheekbone, cheek, nose tip, chin and neck) and body locations (the back of the hand, inner forearm, outer forearm and fingertip). The colour measuring methods included a tele-spectroradiometer (TSR), a spectrophotometer (SP) and a set of skin colour chart used as a visual aid. Also, a characterised digital camera controlled by an imaging system was used to collect facial images. Before the data collection, the short-term repeatability of different settings of the TSR and the SP on measuring human skin colour *in vivo* was determined. And this was used to settle the measurement protocols of the two instruments. The LLSC database was later used to investigate the skin colour distribution between ethnic groups, between genders, between measuring methods. A skin whiteness and blackness scales based on the CIELAB L* and C_{ab}* scales in CIELAB was developed by referencing the vividness and depth formulae, which was developed by Berns (2000). It was found that these scales and

CIELAB hue angle can describe well the property of skin colour of each ethnic group.

Experiment 2 was to investigate the impact of the skin colour on the facial impression attributes. Based on the LLSC database, the gamut of skin colour was defined. Twenty-three attributes used to describe facial skin colours were accumulated. They were classified into two groups (appearance and impression). Two experiments were carried out on a monitor to understand the impact of the skin colour on the perceived facial impression attributes. The first experiment (Experiment 2.1) was to study the relationship between different attributes by 10 observers. The results showed that only four dimensions were required to describe skin facial colours, which were named *Likeable, Sociable, Feminine* and *Youth.* The health was also selected because the traditional Chinese medicine has interested in it. The second experiment (Experiment 2.2) was to scale facial images selected from two ethnic groups and both genders by using these five impression attributes by 24 Chinese observers. The experimental results showed that there were systematic patterns between the impression attributes and the whiteness and hue angle scales. There are some differences between these images for each impression. The ethnic group had an impact on the judgement, but the difference between the Oriental and Caucasian female images was limited. Finally, mathematical models were successfully developed to predict the impressions from the skin whiteness and hue angle data.

Content

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

During human evolution, humans have constantly tried to explore and to understand the universe and the mystery of themselves. One of the topics that have generated interest is how humans perceived impressions vary when seeing different faces. Skin colour may be one of the most conspicuous features of the human face and is important for product development, particularly in the cosmetic and imaging industries. For the cosmetic industry, apart from the formulation of the product, the colour of makeup also has a great impact on the product. This is particularly true for the cosmetic foundation, which was used to unify the skin colour. A study showed that 78% of British women will change their cosmetic foundation if its colour does not match their skin colour (Adby, 2012; Aslam, 2006; Li *et al*., 2008b). A comprehensive skin colour database allows them to understand the colour distribution and colour gamut for different skin groups. So, the cosmetic foundation with matching facial skin colour can be developed. For the imaging industries, including manufacturing of the display, camera, and mobile phone, the quality of skin colour reproduction is utmost important. A comprehensive skin colour database can be used to determine the accuracy of the colour reproduction.

Many skin colour databases were built to determine the range of the skin colour and to assist the skin colour reproduction, such as SOCS database and KAWASAKI skin colour database (see Section 2.2.2). The tele-spectroradiometer (TSR) and the spectrophotometer (SP) were typically used to measure skin colour in these databases. These instruments are commonly used to measure the colour of the samples that have flat geometry and that are spatially uniform in colours, such as textiles, plastics and paper. Human skin is a multilayer soft tissue with colour unevenly distributed on the non-flat surface. Also, it is sensitive to the pressure, temperature and stretch. These properties of the skin increase the uncertainties of the measurement. So, the

measurement results included in these previous databases were included the uncertain variations that were caused by the measurement methods. The earlier databases were accumulated via different instruments, including different ethnicities and different locations so that the data of these databases cannot be used to determine the difference different measurement methods and measurement locations.

The impact of the skin colours on judging perceived facial impression attributes was also extensively studied in the past. Most of the researchers investigated the impact of the skin colour on attractiveness, youth and health in their previous studies. These researchers found that the redness/yellowness of the skin colour can influence the perceived impressions, such as attractiveness and health (Stephen *et al*., 2011; Stephen *et al*., 2009b; Stephen *et al*., 2009a; Stephen *et al*., 2012). However, there was a limited range of skin colours and a small set of perceived impression attributes that were examined in many of these studies. The preferred skin colour was also studied by many researchers. For different geographical areas and cultures, the preferred skin colour could vary. For example, Europeans preferred tanned skin colour which is associated with being healthy and attractive (Picton, 2013). For East Asia, a whiter skin colour is preferred, as the tanned skin is associated with working in a tough environment (Li *et al*., 2008a). For different genders, the preferred skin colour can also be different (Aoki, 2002). But, limited cross-culture study was carried out.

1.2 Aim and objectives

The overall goal of the present study is to investigate the influence of skin colour on the judgement of the perceived facial impression attributes. To achieve this goal, two objectives are set.

The first objective is to establish a comprehensive skin colour database. This skin colour database will include the skin colour of a large number of subjects from different ethnicities, and different instruments will be used to measure body and facial locations for each subject. The specific tasks included to:

- Investigate the short-term repeatability of each instrument for measuring skin colours in vivo,
- Understand the distribution of skin colours between different locations of a subject, between different ethnic groups, between genders and between instruments
- Formulate the relationship between skin colours and colorimetric data.

The second objective is to study the impact of the skin colour on the perceived facial impression attributes that were frequently and popularly used. The tasks include to:

- Accumulate perceived facial impression attributes,
- Conduct psychophysical experiment to find the basic dimensions to describe facial impression,
- Model the relationship between the colorimetric values and the perceived facial impression attributes for each ethnic and gender group.

1.3 Thesis outline

Six chapters are included in this thesis. The current chapter is the Chapter 1, which describes the background and goals, objectives and tasks of the present study. Chapter 2 reviews colour theories and the previous studies related to the present study. Chapter 3 introduces the experimental design

and preparation of the present study. The instrumental evaluations, short-term repeatability tests and characterization modelling accuracy tests are also included in this chapter. Chapter 4 focuses on an experiment to collect skin colour data (Experiment 1). The distributions of the skin colour of different ethnicities and of different locations are reported and studied in this chapter. Based on the skin colour database, the skin whiteness and skin blackness scales were proposed. Chapter 5 reports the results from the two psychophysical experiments (Experiment 2.1 and 2.2) which investigate the impact of skin colour on perceived facial impression attributes. The range of skin colour examined is determined through the skin colour database that was described in Experiment 1. The influence of skin whiteness and hue angle on the perceived facial impression attributes is also described in this chapter. Also, the influence of the gender of the observer, and the gender and ethnicity of the subject in the stimuli image on the judgements is investigated and is reported in this chapter. The sixth chapter is the conclusions of the present study.

Chapter 2 Literature review

In this chapter, six sections were included. The first section reviewed the literature of the colorimetry topics, including the CIE recommended illuminants, measurement geometry, tristimulus values and uniform colour space. Two colour measuring instruments were also introduced in this section. The second section reviewed the skin colour measurement methods, skin colour databases and the previous studies of the skin colour and facial impressions. The third section reviewed the perceived whiteness and blackness related scales that proposed in the previous study. This section provides background information for the skin whiteness and skin blackness scales that were proposed in the present study. The fourth section reviewed the characterisation models used to characterise the imaging devices (digital camera and display) in the present study. The psychophysical methods that used in the present study were then introduced in the fifth section. Subsequently, the previous research related to the skin colour measurement, database and perceived impressions were reviewed. Finally, the statistical methods that used later to analyse the experimental results were reviewed.

2.1 Colorimetry

The International Commission on Illumination (CIE) recommends standards to measure colour, to describe the colour and to enable colour communication between different media. The CIE standard illumination, CIE recommend measuring geometries, CIE colour spaces and CIE colour difference formulae that used to evaluate the colour measurements from the present study were reviewed in this section. Apart from the CIE recommendations, two colour measuring instruments that were used to measure skin colour were also introduced. Two colour order systems that widely used were also reviewed.

2.1.1 CIE Standard illuminants

The light source is essential for observing the colour of a reflective specimen. The optical radiation that emits from a light source can be described by using the spectral power distribution (SPD) which can be defined as the distribution of energy at each wavelength, typically from 380 to 780 nm in the visible spectrum (CIE, 2011). The relative SPD is normalised SPD at 560 nm equal to 100. The colour temperature is defined as the temperature of the blackbody radiator or Planckian radiator which had the same chromaticity as the given stimulus (CIE, 2011). The Planckian locus is the locus illustrates the change of the chromaticity of the radiator with different temperature (CIE, 2011). For some of the light source, their chromaticities do not lie on the Planckian locus. In such cases, the correlated colour temperature is used. The correlated colour temperature (CCT) is defined as "*temperature of the Planckian radiator having the chromaticity nearest the chromaticity associated with the given spectral distribution on a diagram where the (CIE 1931 standard observer based) (u',2/3v') coordinates of the Planckian locus and the test stimulus are depicted*" (CIE, 2011). Luminance, as defined in the Oxford dictionary, is "*the luminous intensity per unit area when the light travels in a given direction"* (Dictionary, 2007). The International System of Units (SI) of luminance is candela per square meter. It can be used to describe the brightness of the light source and of the monitor in the display industry.

CIE recommended illuminants to describe various of light sources (CIE, 2006), including CIE illuminant A and CIE illuminant D. The CIE illuminant A is "*intended to represent typical, domestic, tungsten-filament lighting. Its relative spectral power distribution is that of a Planckian radiator at a temperature of approximately 2856 K.*" (CIE, 2006). The CIE illuminant D includes series of illuminants that used to represent the different phase of daylight which are illuminants D50, D55, D65 and D75 for the correlated colour temperature of 5000K, 5500K, 6500K and 7500K. Figure 2.1.1.1 shows the relative SPD of CIE illuminant D series.

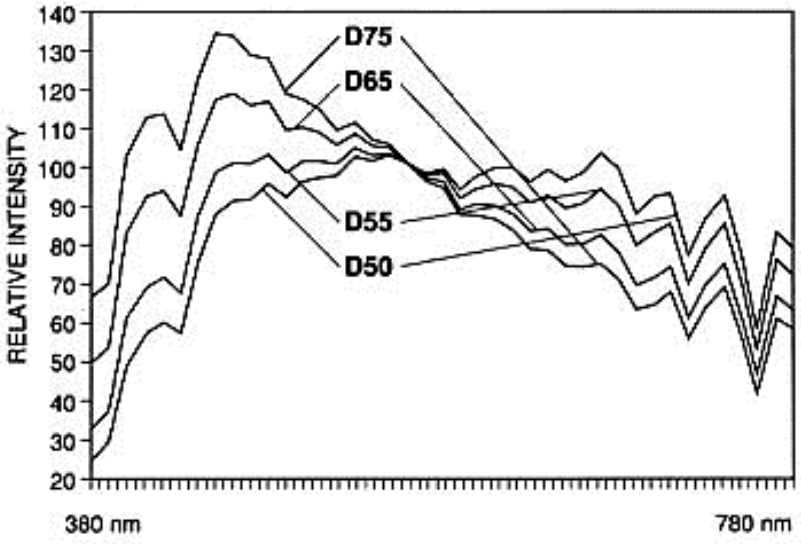


Figure 2.1.1.1 The relative spectral power distribution of CIE illuminant D (Sersen, 1989).

2.1.2 Colour rendering index

The colour rendering index (CRI) is used to evaluate the colour rending property of light sources (CIE, 1995a; CIE, 2011). The CIE defined it as "*the effect of an illuminant on the colour appearance of objects by conscious or subconscious comparison with their colour appearance under a reference illuminant*" (CIE, 1995a; CIE, 1987). The CIE (1965) defined colour rendering index R_a, which can be calculated by Equation 2.1 (CIE, 1995a).

$$R_a = 100 - \frac{4.6}{8}\sum_{i=1}^{n} \Delta E_i \qquad \textbf{Equation 2.1}$$

where ΔE_i is the colour difference between the sample measured under the ideal CIE standard illuminant simulator and the test light sources in CIE 1964 W*U*V* colour space (CIE, 1995a). n is the number of test samples.

For the formula proposed in 1964, eight Munsell colour samples are used to calculate CIE W*U*V* colour difference, as these samples have low saturation

and cover the whole range of hue. The spectral reflectances of these eight test colour samples are shown in Figure 2.1.2.1. Six supplementary test colours were added in 1995 (CIE, 1995a). Four of the supplementary colours were highly chromatic red, yellow, green and blue. The other two were colours of human skin and foliage (CIE, 1995a). A colour rendering index equal to 100 indicates that the test colour samples have no difference under the standard illuminant and test light source. And the test light source can then be said to be perfect at colour rendering.

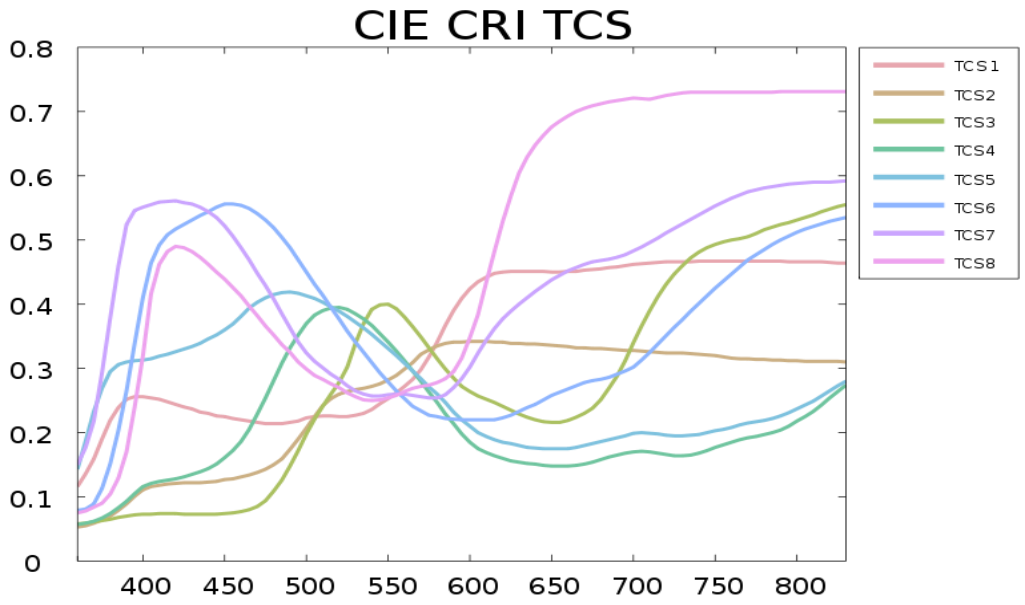


Figure 2.1.2.1 The test colour samples for colour quality scale. (Adoniscik, 2008a).

2.1.3 CIE measuring geometry

CIE recommended standard geometries for measuring spectral radiation reflected from opaque specimens (CIE, 2004). These geometries included 45°x:0°, 0°:45°x, 45°a:0°, 0°:45°a, di:8°, 8°:di, de:8° and 8°:de. In these geometry symbols, the *x* of the *45°x* means the incident beam light is used. The *a* of the *45°a* means a ring light is used. The *i* of the *di* means the gloss trap is included. The e of the *de* means the gloss trap is excluded. The 45°x:0°,

0°:45°x, 45°a:0° and 0°:45°a can be used when TSR was the measurement instrument. The di:8°, 8°:di, de:8° and 8°:de can be used when using the instrument with integrating sphere built inside, such as the SP. Figure 2.1.3.1 illustrates the 45°x:0°, 0°:45°x, di:8° and 8°: di geometries.

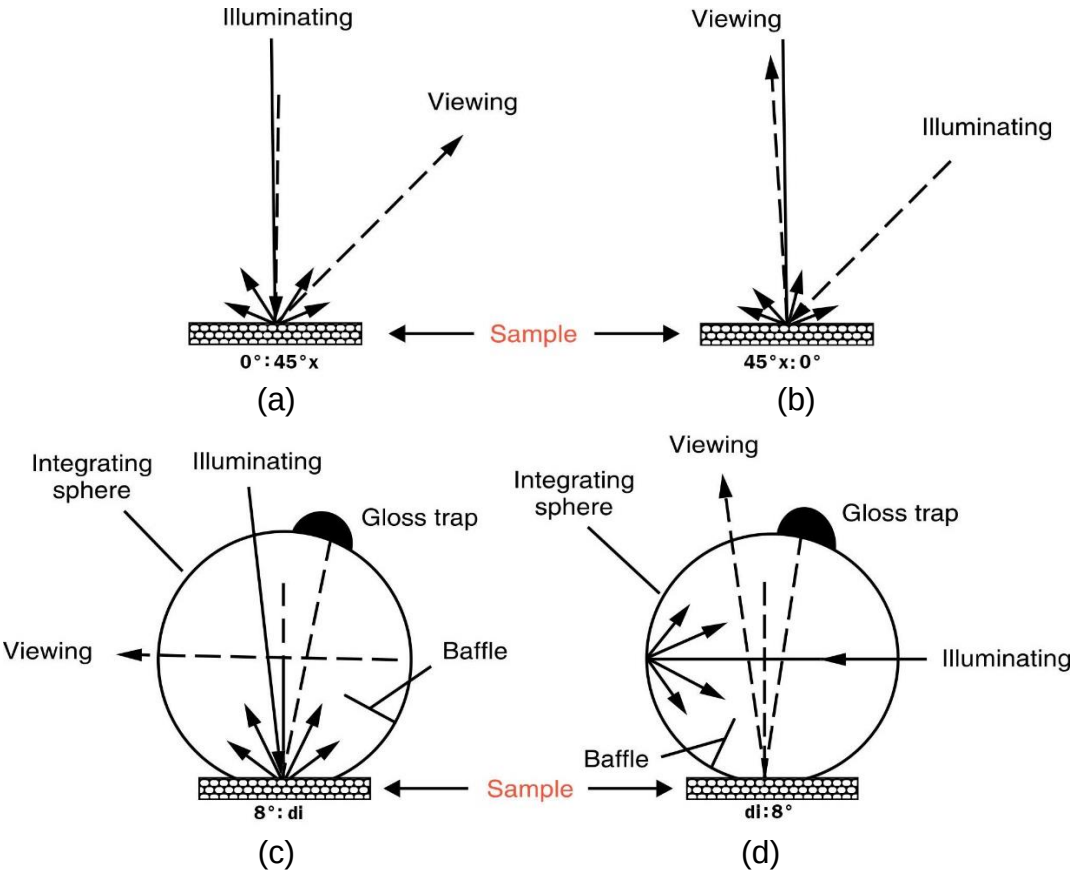


Figure 2.1.3.1 Four of CIE recommended measurement geometries (reflective sample) (Hunt and Pointer, 2011, P104) (a) 0°:45°x; (b) 45°x:0°; (c) 8°:di; (d) di:8°.

As shown in the Figure 2.1.3.1 (a), the 0°:45°x geometry is to have the light source illuminated the specimen at the direction perpendicular to the specimen. The measuring angle was defined 45°±2° to the direction perpendicular to the specimen. The 45°x:0° geometry swaps the direction of illumination and measurement. For the di:8°, 8°:di, de:8° and 8°:de, the 8°:di and 8°:de, the direction of the emitted light is no more than 10º to the direction

perpendicular to the specimen. It can be seen from Figure 2.1.3.1 (c) and (d) that the gloss trap or specular port is included the integrating sphere for the inclusion and exclusion of the specular reflection from the glossy sample.

2.1.4 CIE Standard Colorimetric Observer

The trichromatic matching experiment was to investigate the visual response of the human eye over the visible spectrum based on the trichromatic theory of colour vision (CIE, 2014). Two experiments, which carried out by Guild (1931) and Wright (1969), were used to investigate these responses. These experiments used 2˚ viewing angle. Figure 2.1.4.1 illustrates the experimental setup.

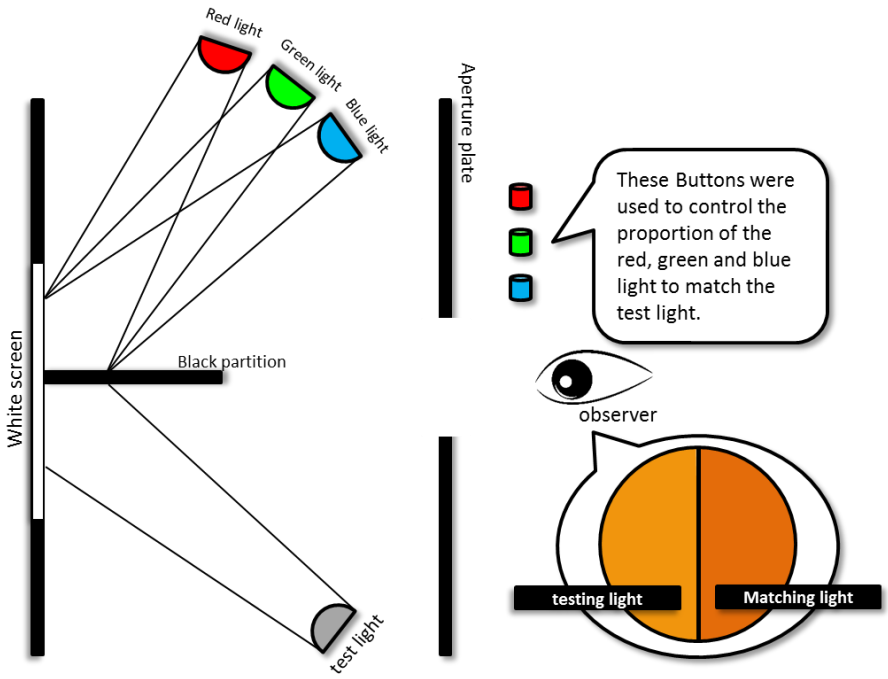
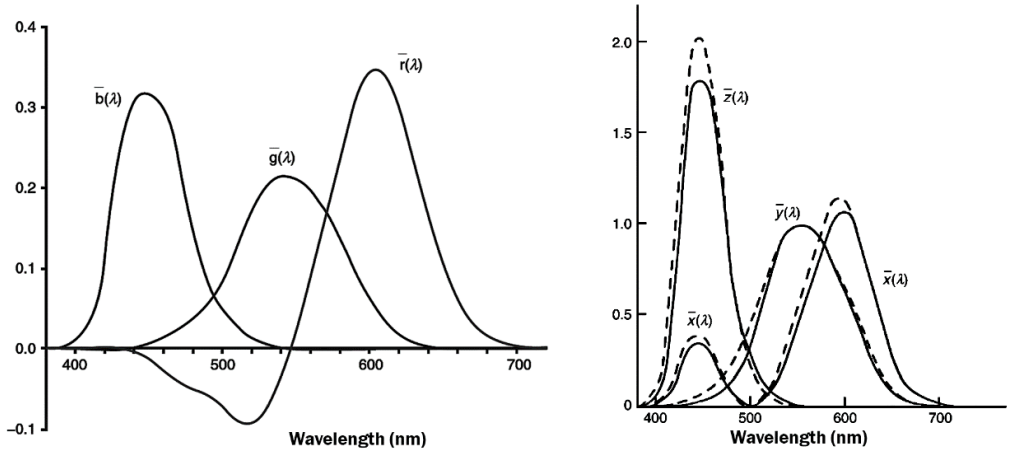


Figure 2.1.4.1 The set-up of the trichromatic matching experiment.

The experimental results were transformed mathematically into the red light at 700 nm, the green light at 546.1 nm and the blue light at 435.8 nm, to match the test light (CIE, 2004). The measurement results are shown in Figure

2.1.4.2 (a), where $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ represent the amount of R, G and B needed to match a stimulus at each wavelength and named *CIE 1931 RGB colour specification system* (CIE, 2004). The problem of the CIE RGB colour specification system was the negative response from $\bar{r}(\lambda)$, which brings complexity into the calculation. To reduce the complexity, a CIE colour matching function introduced, as shown in Figure 2.1.4.2 (b). The negative value from $\bar{r}(\lambda)$ transformed into positive values. And the R, G and B symbols were transformed into $\bar{x}$, $\bar{y}$ and $\bar{z}$ to avoid confusion. It named *CIE 1931 colour-matching function*. The 10˚ system, CIE 1964 colour matching functions ($\bar{x}_{10}$, $\bar{y}_{10}$, $\bar{z}_{10}$)were then recommended by CIE for the observing or measuring field size greater than 4 degrees (CIE, 2004).



(a) (b)

Figure 2.1.4.2 The colour-matching functions. (a) the CIE 1931 RGB colour specification system (Hunt and Pointer, 2011, P30); (b) CIE 1931 colour-matching function (normal line) and CIE 1964 colour-matching function (dashed line) (Hunt and Pointer, 2011, P34).

2.1.5 CIE colour space

2.1.5.1 CIE Tristimulus Values (XYZ)

The CIE tristimulus values (XYZ) of a surface can be achieved through integrating the reflectance of the object, the SPD of the CIE standard illuminant and CIE colour matching functions. The formulae are given in Equation 2.2.

$$X = k \sum_{\lambda=380}^{780} \mathrm{S}(\lambda)\mathrm{R}(\lambda)\bar{x}(\lambda)d\lambda$$
$$Y = k \sum_{\lambda=380}^{780} \mathrm{S}(\lambda)\mathrm{R}(\lambda)\bar{y}(\lambda)d\lambda \qquad \textbf{Equation 2.2}$$
$$Z = k \sum_{\lambda=380}^{780} \mathrm{S}(\lambda)\mathrm{R}(\lambda)\bar{z}(\lambda)d\lambda$$

where

$$k = 100/\sum_{\lambda=380}^{780} \mathrm{S}(\lambda)\mathrm{R}(\lambda)\bar{z}(\lambda)d\lambda$$

In these formulae, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the CIE colour matching functions. $R(\lambda)$ is the spectral reflectance of the specimen. $\mathrm{S}(\lambda)$ is the spectral distribution of the light source that be used in the colour measurement. For a perfect white diffuser, the value of the Y is assigned into 100.

To present the relative magnitudes of tristimulus values (X, Y and Z), the chromaticity coordinates (x, y and z) are introduced. The chromaticity coordinates can be calculated via Equation 2.3.

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z}$$
$$z = \frac{Z}{X+Y+Z}$$

Equation 2.3

where $x + y + z = 1$

In 1931, a two-dimensional diagram, the CIE x, y chromaticity diagram, was proposed to specify colours. The CIE 1931 x, y chromaticity diagram is shown in Figure 2.1.5.1. However, the colour distribution of this diagram is not uniform. The distance between two colours in this diagram cannot be used to accurately define the perceived colour difference. A uniform chromaticity scale diagram, named the CIE 1976 uniform chromaticity scale diagram, was introduced (see next section).

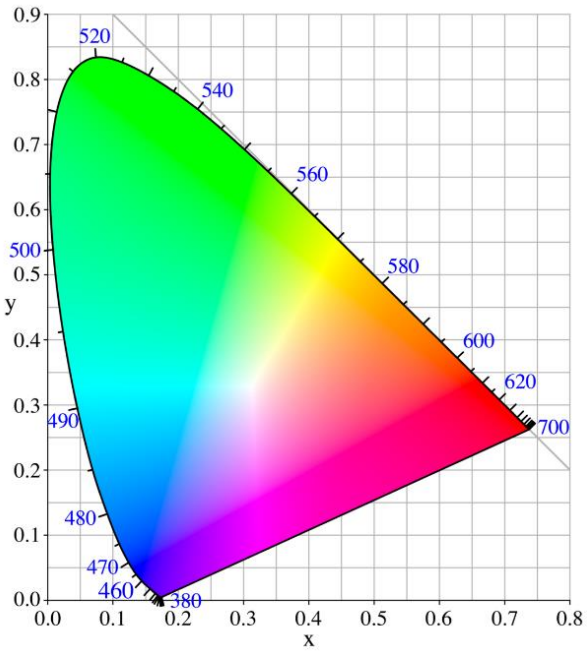


Figure 2.1.5.1 CIE 1931 x, y chromaticity diagram (BenRG, 2009).

2.1.5.2 Uniform colour space

As mentioned in the previous section, the CIE1931 x, y chromaticity diagram is not visually uniform. So, the CIE 1976 uniform chromaticity scale diagram or CIE 1976 UCS diagram, which has better uniformity, was recommended, as shown in Figure 2.1.5.2. This chromaticity diagram was sometimes used to specify the colour property of the monitor in the past research (Gong *et al*., 2012; Berns, 1996b), such as colour gamut of the display. The u'v' can be achieved via the known xy or XYZ via Equation 2.4.

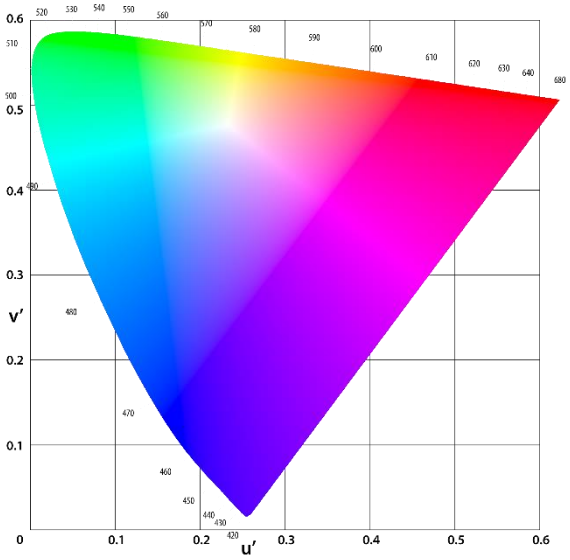


Figure 2.1.5.2 The CIE 1976 uniform chromaticity scale diagram or CIE 1976 UCS diagram (Adoniscik, 2008b).

$$u' = \frac{4x}{-2x + y + 3} = \frac{4X}{X + 15Y + 3Z}$$

$$v' = \frac{9y}{-2 + 12y + 3} = \frac{9Y}{X + 15Y + 3Z}$$

Equation 2.4

Even though this chromaticity diagram has better uniformity, the shortcomings of the chromaticity diagram were not overcome. The shortcoming of the

chromaticity diagram was the chromaticity diagram only can be used to compare the colour with same luminance (Y) and luminance factor (Y/Y_n) (Hunt and Pointer, 2011, p49-50). Therefore, two uniform colour spaces were recommended by CIE, the CIELUV and CIELAB colour space, to cross compare the colour with different luminance and luminance factor. The CIELUV or CIE $L^*u^*v^*$ colour space was recommended to be used for describing the colour of the self-luminous objects, such as a monitor. The CIELAB or CIE $L^*a^*b^*$ colour space was recommended to use for describing the colour of the object surface (Malacara, 2003), such as textile, paper and skin colour. Equation 2.5 are the formulae used to transform CIE 1931 XYZ into CIELAB, as listed below (ISO, 2007).

$$L^* = 116f(Y/Y_n) - 16$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \qquad \textbf{Equation 2.5}$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)]$$

where

$f(X/X_n) = (X/X_n)^{1/3}$ for $(X/X_n) > (6/29)^3$

$f(X/X_n) = (841/108)(X/X_n)^{1/3} + 4/29$ for $(X/X_n) \leq (6/29)^3$

$f(Y/Y_n) = (Y/Y_n)^{1/3}$ for $(Y/Y_n) > (6/29)^3$

$f(Y/Y_n) = (841/108)(Y/Y_n)^{1/3} + 4/29$ for $(Y/Y_n) \leq (6/29)^3$

$f(Z/Z_n) = (Z/Z_n)^{1/3}$ for $(Z/Z_n) > (6/29)^3$

$f(Z/Z_n) = (841/108)(Z/Z_n)^{1/3} + 4/29$ for $(Z/Z_n) \leq (6/29)^3$

where X_n, Y_n and Z_n are the tristimulus values of the reference white.

The CIELAB chroma (C^*_{ab}) and the CIELAB hue angle (h_{ab}) can be calculated via the Equation 2.6 and Equation 2.7, respectively.

$$C^*_{ab} = ({a^*}^2 + {b^*}^2)^{1/2} \qquad \textbf{Equation 2.6}$$

$$h_{ab} = \arctan(b^*/a^*) \qquad \textbf{Equation 2.7}$$

Hue is a visual perception which appeared to be similar to one proportion, or to two proportions of the perceived colours red, yellow, green, and blue. Chroma is the colourfulness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white or highly transmitting (CIE, 2011).

The a^* and b^* also can be calculated from C^*_{ab} and h_{ab}, as shown in Equation 2.8.

$$a^* = C^*_{ab} \sin(h_{ab})$$

$$b^* = C^*_{ab} \cos(h_{ab}) \qquad \textbf{Equation 2.8}$$

The L^*, C^*_{ab} and h_{ab} are similar to the Munsell Value, Munsell Chroma and Munsell Hue colour coordinates, respectively. Many skin colour charts were developed by referencing the Munsell colour system. The Munsell colour system will be further discussed in Section 2.1.8.1.

2.1.6 CIELAB colour difference formula

The colour difference between two colours in the CIE$L^*a^*b^*$ colour space can be calculated through CIE $L^*a^*b^*$ colour difference or CIELAB colour difference formula, as shown in Equation 2.9.

$$\Delta E^*_{ab} = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$$

Or $\Delta E^*_{ab} = [(\Delta L^*)^2 + (\Delta H^*_{ab})^2 + (\Delta C^*_{ab})^2]^{1/2}$ **Equation 2.9**

where

$$\Delta H^*_{ab} = 2(C^*_{ab,1} - C^*_{ab,2})^{1/2} \sin(\Delta h_{ab}/2)$$

DH^*_{ab} is the CIELAB hue difference. $C^*_{ab,1}$ and $C^*_{ab,2}$ in the formula are the CIELAB chroma values of two samples.

After CIELAB colour difference formula, some other colour difference formulae were proposed, such as CMC (1:c) (Clarke, McDonald and Rigg,1984), CIE94 colour difference formula (McDonald and Smith, 1995), and CIEDE2000 (Luo, Cui and Rigg, 2001). CMC (1:c) formula was developed for the colorant industry to calculate the small colour difference (Clarke *et al*., 1984; Hunt, 1986; Standard, 1988). After CIE $L^*a^*b^*$colour difference formula, CIE recommend CIE94 (CIE, 1995b). This formula was soon replaced by CIEDE2000. CIEDE2000 formula developed based on CIE94 and mainly refined the colour difference calculation in the blue region (CIE, 2001; Nobbs, 2002; Luo, Cui and Rigg, 2001). In the past studies, CIE $L^*a^*b^*$ colour difference formula was wide choice to use to calculate skin colour difference (Zeng and Luo, 2009; Zeng and Luo, 2013; Stephen *et al*., 2009a; Stephen *et al*., 2009b; Stephen, Coetzee and Perrett, 2011). In the present study, the CIE $L^*a^*b^*$ colour difference formula was also chosen to evaluate colour differences.

2.1.7 Colour measuring instrument

The colour of the surface can be measured by many methods. One of the methods is the visual assessment. This method requires observers with normal vision to select samples from a colour chart which matches the colour of the target sample under a standard illuminant simulator (such as D65). This

method is based on subjective judgment, but other measurement results often closely match observers' observations (Hunt, 1986). Comparing to the colorimeter, visual assessment has low repeatability but it is inexpensive and easy to operate for a non-colour expert. For the cosmetic and dental industry, colour matching is still the major method to measure and evaluate the colour of teeth or skin complexion (see Section 2.2.1).

Spectrophotometers and spectroradiometers were two types of instruments that can be used to measure the spectral reflectance or the SPD of the object surface. These instruments have better repeatability than the visual assessment and widely used in colour study (Hunt and Pointer, 2011, p99; Hunter and Harold, 1987). Figure 2.1.7.1 and Figure 2.1.7.2 show the basic features of a spectrophotometer and a spectroradiometer, respectively.

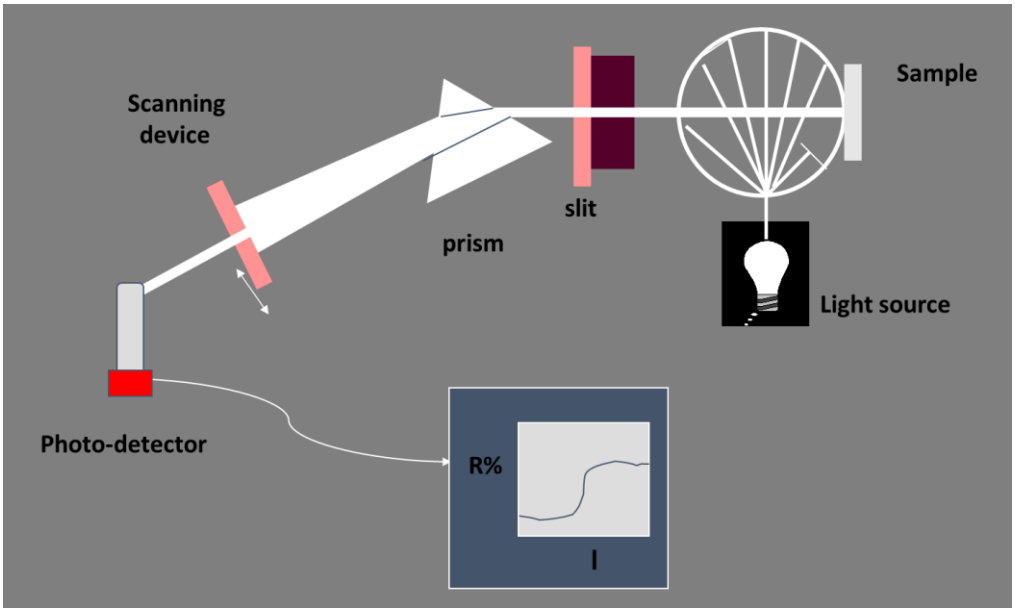


Figure 2.1.7.1 The basic features of a spectrophotometer (Luo and Rhodes, 1999).

The spectrophotometer is *" an instrument for measuring the ratio of two values of a radiometric quantity at the same wavelength*" (CIE, 2011). This can be understood as it is the instrument which gained the spectral reflectance of the

sample through comparing the spectrum reflected from the sample and the spectrum reflected from a reference white (Hunt and Pointer, 2011, p101; Berns, 2000, p88). As shown in the Figure 2.1.7.1, the spectrophotometer includes a light source, an integrating sphere, and a photo-detector. The spectrum that reflected from the sample is detected by the photo-detector and then compared with the spectrum that reflected from a reference white to gain the reflectance of the sample.

Ideally, the perfect reflecting diffuser is chosen to be the reference white. The perfect reflecting diffuser is defined as "*diffuser exhibiting isotropic diffuse reflection with a reflectance equal to unity*", which is not available for practical measurement (Hunt and Pointer, 2011, p102; CIE, 2011). But it can be measured through calibrating a working standard white tile which normally provided by the instrument manufacturer.

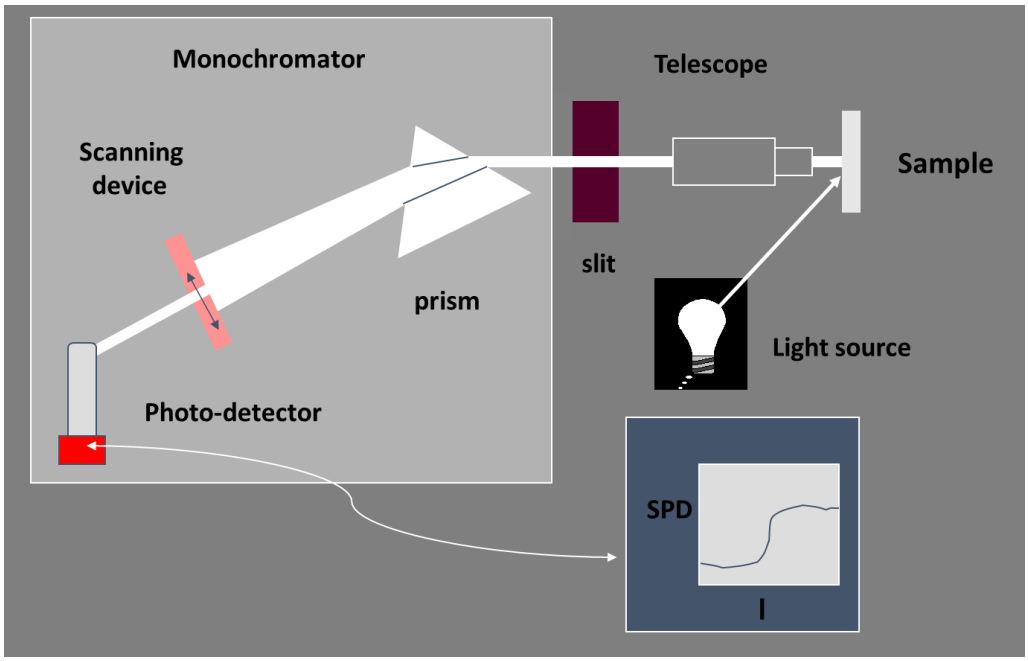


Figure 2.1.7.2 The optical elements of a spectroradiometer (Luo and Rhodes, 1999).

The spectroradiometer is "*an instrument for measuring radiometric quantities in narrow wavelength intervals over a given spectral region*" (CIE, 2011). This instrument can be used to simulate the human observing an object under certain conditions (Berns, 2000, p92; Hunt and Pointer, 2011,p100). As shown in Figure 2.1.7.2, the spectroradiometer includes a telescope and a monochromator. For measuring a reflective sample, a light source is needed. The reflected spectrum of the sample goes through the telescope and then is detected by the monochromator. The SPD can be obtained from a spectroradiometer. The SPD can be transformed into reflectance by dividing the SPD of a reference white which measured under the same light source. Then the reflectance can be transformed into CIEXYZ tristimulus values and then can be transformed into the coordinates of CIELAB (see Equation 2.2 and Equation 2.5).

2.1.8 Colour order system

Many colour order systems were developed and used in different applications. Two of the most popular colour order systems that related to the present study are described in this section. These colour systems are the Munsell colour system and the Natural Colour System. For Munsell colour system, many skin colour charts, such as the *Skin Color Chart*® (*L'Oréal*®) and *Pantone SkinTone*™ *Guide*, were developed via referencing this colour system. For the Natural colour system, the vividness and depth formula that developed by Berns (2014) were referenced this system.

2.1.8.1 Munsell colour system

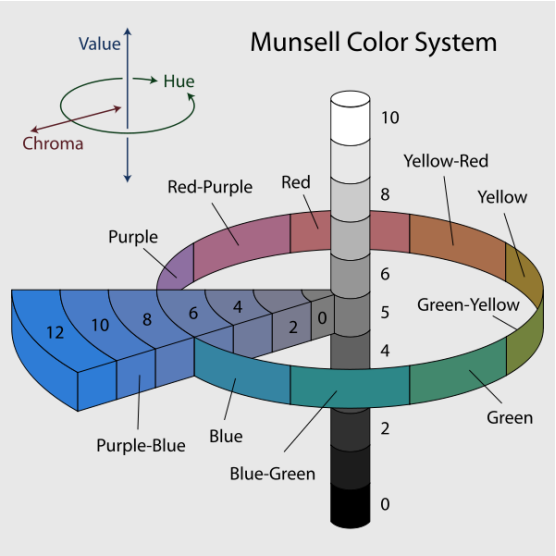
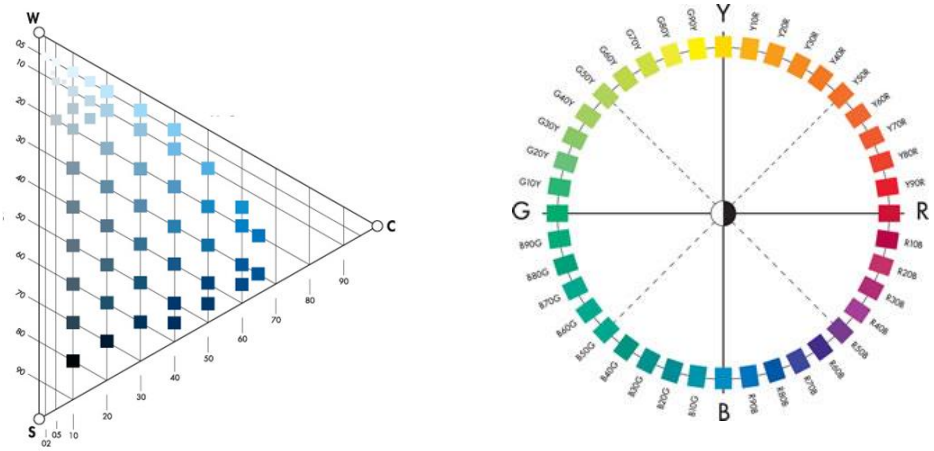


Figure 2.1.8.1 The Munsell Colour System (Jacobolus, 2007).

The Munsell colour order system uses Munsell Value, Munsell Chroma and Munsell Hue to specify colours (Cooper, 1941). Figure 2.1.8.1 illustrates the Munsell colour system. The Munsell Value is similar to the CIE lightness which ranges from 0 to 10 for black to absolute white. The Munsell Hue contains 5 principal hues (5R(red), 5Y(yellow), 5G (green), 5B (blue) and 5P (purple)) and they can be further divided into 5 intermediate hues (5YR, 5GR, 5BG, 5PB and 5RP). Munsell Chroma is *"the departure degree of a color from the neutral color of the same value"* (Rhodes, 2002). The range of the Munsell Chroma starts from 0 (neutral) and extended out to the maximum at each hue.

2.1.8.2 Natural colour system



(a) (b)

Figure 2.1.8.2 The Natural Colour System (a) NCS colour triangle; (b) NCS colour circle (NCS, 2017).

The Natural Colour System (NCS) was promoted by Hering in mid-1870 and it is widely used by paint industry, designer and architects (Berns, 2000, p39; Kuehni and Schwarz, 2008,p109-110). In the NCS system, six elementary colours (red (R), yellow (Y), green (G), blue (B), white (W) and black (S)) are included in this colour system. Three scales are used to describe a colour, including hue, blackness and chromaticness, and they are expressed in percentage. The NCS colour triangle and NCS colour circle are two diagrams that used to locate a colour in the NCS system, as shown in Figure 2.1.8.2 (NCS, 2017). The NCS colour circle can be used as a reference to the hue scale. The value of the blackness and chromaticness scale can be indicated via the NCS colour triangle. In the NCS colour triangle, three variables are included, white (W), black (S) and chromaticness (c). The NCS system defined that the sum of the W, S and c is equal to 100 (NCS, 2017). The W and S are the two neutral elementary colours of the NCS system. The c represents the colour with the maximum chromatic of the colours with the same hue. This colour system was claimed to give a more natural description by using scales like whiteness and blackness and it was easier to be

understood than the lightness or Munsell value for the architects (Whitfield *et al*., 1986; Gloag and Gold, 1978; Hunt and Pointer, 2011, p170). The chromaticness and blackness are claimed to relate to some terms that designers use to describe the colour, such as *weight* and *nuance* (Hunt and Pointer, 2011, p170).

2.2 Skin colour and facial impression attributes

Skin can be considered as the largest human organ and a vital facial feature that characterises the human face. The colour of the skin reflects ethnic, habitat area and health (Alam, 2010). Studies were found that the skin colour can be influenced by the pigments in the skin, such as melanin, erythema, haemoglobin and carotene (Edwards and Duntley, 1939; Stamatas *et al*., 2004). It also can influence the impression perceived by others and the preference of the facial images (Hill, 2002; Stephen *et al*., 2009b; Van den Berghe and Frost, 1986; Yuan *et al*., 2011; Zeng and Luo, 2013). To study the skin colour, a skin colour database which can be used to determine the skin colour range and used for skin colour reproduction is essential. To build a skin colour database, the measuring method is crucial. In this section, the previous skin colour databases and different skin colour measuring methods are reviewed. The previous studies which were investigated the impact of the facial features on the perceived impressions are reviewed.

2.2.1 Skin colour measurement

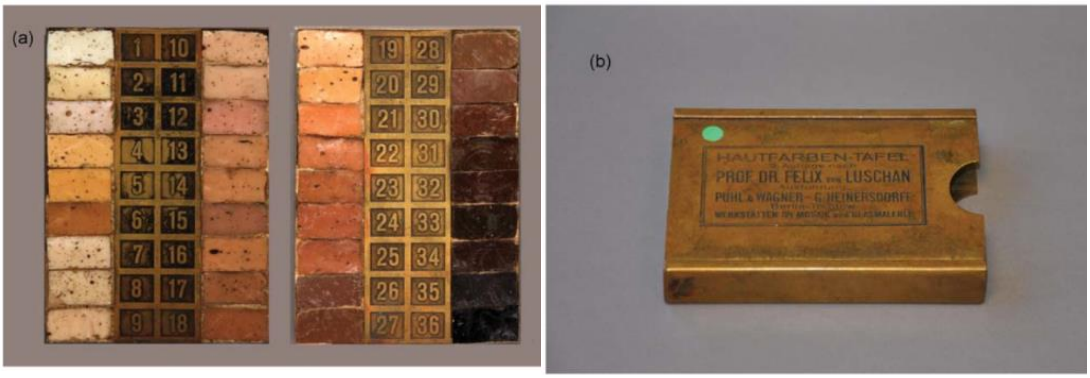
Unlike paper and plastic, human skin is a biological material and it is a multi-layer soft tissue with non-uniform colour distributed on the surface. The geometry of the skin on the human body is not flat (Muehlenbein, 2010). These skin properties increase the uncertainty of measuring skin colour *in vivo*.

TSR, SP and colour charts are the tools that are typically used to measure skin colour for study or for evaluating dermatology treatment.

2.2.1.1 The skin colour chart

The colour chart is the tool which used to measure skin colour before the colour measurement instruments were introduced. *Von Luschan's chromatic scale* or *Von Luschan's scale* is a skin colour chart that has been widely used. Figure 2.2.1.1 shows this colour chart. This colour chart contains 36 coloured opaque glass tiles (Swiatoniowski *et al*., 2013; Broca, 1879; Robins, 2005). This skin colour chart is still be used to measure skin colour and evaluate skin treatments in the recent studies (Mattos *et al*., 2016; Zhdanova *et al*., 2014; Treesirichod *et al*., 2014). But these 36 glass tiles are limited for determining the skin colour with a wider range (Motokawa *et al*., 2006).

The *Fitzpatrick skin prototype scale* was developed to classify human skin colour to predict UV influence on the skin (Fitzpatrick, 1988). This scale classified skin colour into 6 types, from I to VI, based on the original colour of the skin, hair and eye. Type I was marked as the skin that is most easily sunburnt and Type VI was marked skin that does not get sunburnt at all. This scale was used as a reliable guide to select suitable UV protecting methods for preventing skin cancer (Fasugba *et al*., 2014; Agency, 2017; Magin *et al*., 2012). This skin colour scale and *Von Luschan's scale* can be used together to classify skin colour for UV protection (Mattos *et al*., 2016).



(a) (b)

Figure 2.2.1.1 Von Luschan glass tiles (a) the 36 opaque glass tiles (b) the outer packing of the glass tiles (Swiatoniowski *et al*., 2013).

The Taylor hyperpigmentation scale is a skin colour chart that was developed for evaluating hyperpigmentation treatment. This colour chart contains 15 coloured plastic cards that were designed based on the *Fitzpatrick skin prototype* Types I to VI. Ten bands on each card represent an increasing grade of hyperpigmentation, as shown in Figure 2.2.1.2 (Taylor *et al*., 2005). Firstly, the clinician needs to find the card match the normal skin colour of the subject. Then, the clinician used the ten bands on the selected chart to evaluate the hyperpigmentation area. This colour chart has been shown to be reliable for evaluating hyperpigment treatment (Taylor *et al*., 2006).

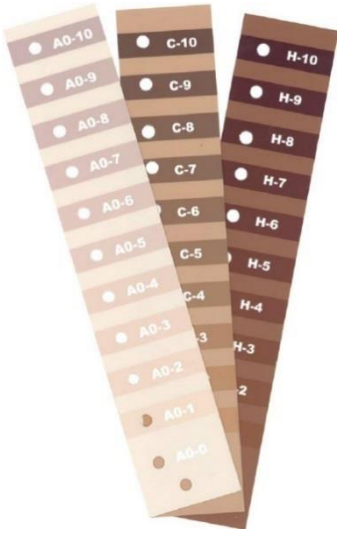


Figure 2.2.1.2 Taylor hyperpigmentation scale (Taylor *et al*., 2006).

The skin colour charts that introduced above were mainly used to satisfy the special needs, such as diagnose dermatology diseases. They may not suitable to measure skin colour, as the range and number of the colours included in these charts are limited. Two skin colour charts that included more skin colours than the others were developed.

The *Skin Color Chart*® is a colour chart developed by *L'Oréal*® for assisting their clinicians in evaluating the skin care products, such as skin whitening creams (De Rigal *et al*., 2007). This colour chart is a printed, paper-based colour chart which contained 52 fan charts, as shown in Figure 2.2.1.3. This skin colour chart was developed based on the skin colour database which data were collected via a *Chromasphere*®. *Chromasphere*® is an illumination system invented by *L'Oréal*®, as shown in Figure 2.2.1.3. This equipment provides uniform illumination to the face of the subject. Four facial locations were measured by two TSRs and more than 1000 subjects were measured. The colour of charts was evenly selected from the gamut of the measurement results in the CIELAB $L^* h_{ab}$plane (De Rigal *et al*., 2007). This skin colour chart was only available within *L'Oréal* ®.

(a) (b)

Figure 2.2.1.3 *L'Oréal* ® developed *Skin Color Chart* ® (a) the *Skin Color Chart* ® ; (b) the *Chromasphere*® (De Rigal *et al*., 2007)**.**

The *Pantone SkinTone™ Guide* is a skin colour chart successfully used in personalised cosmetic shopping (Sephora, 2017). This skin colour chart included 110 fan charts that selected from the real skin colour of more than 1000 subjects which measured by a spectrophotometer (Adby, 2012). This chart is recommended to be used under CIE D65 simulator for the best colour matching result. Each chart of the *Pantone SkinTone™ Guide* has a 10 mm size hole for observing a target colour. The *Pantone SkinTone™ Shades* were arranged to include 110 colours in the *Pantone SkinTone™ Guide* in a lightness and redness/yellowness related plane.

(a) (b)

Figure 2.2.1.4 Skin colour chart developed by Pantone (a) the *Pantone SkinTone™ Guide*; (b) the *Pantone SkinTone™ Shades.*

2.2.1.2 The skin colour measurement instruments

TSR and SP are two types of instruments to measure the spectral information of samples (see Section 2.1.7). The TSR is a non-contact type instrument, which requires an external light source. The measuring field size of the TSR related to the measuring distance and measuring field angle. The SP is a contact type of instrument. Some of them have a built-in integrating sphere and a light source.

These two types of instruments were frequently used to measure skin colour in the previous studies. For example, *L'Oréal*® used tele-spectroradiometers to measure skin colour and built a *Skin Color Chart*® based on the TSR measuring results (De Rigal *et al*., 2007); *Pantone SkinTone™ Guide* was built based on the skin colour data which measured by a spectrophotometer. KAWASAKI skin colour database included the skin colour measured results of both types of instruments.

Both instruments can potentially give a more precise measurement of skin colour than visual assessment. But the difference between these two types of

instrument is uncertain. The KAWASAKI skin colour database included two types of instrument's measurement results at 3 facial locations. Based on their published data, the average colour difference between two instruments was calculated to be about 4 ΔE^*_{ab}. This indicates that two types of instrument can result in different measurement results. With limited measurement information, the factors caused the difference in measuring result were unknown. Not only the measuring different between two instruments, the uncertainty of using these instruments to measure skin colour in vivo also unknown.

As mentioned earlier, the visual assessment, TSR and SP were the methods that be used to measure skin colour. But the study of the comparability between these methods was limited. This issue appeared crucial for studying the human evolutionary biology. As the skin colour data collected from the previous study may not be able to measure again (Jablonski, 2004; Muehlenbein, 2010).

2.2.1.3 The measurement of other human skin properties

Apart from the skin colour, the other properties of the human skin are also widely interested by the dermatology, computer graphics and cosmetic industry. These properties included the degree of melanin and the haemoglobin, surface scattering, translucency.

The melanin and haemoglobin are two components in the human skin that contribute to the human skin colour. For dermatology, these two components also related to the skin disease, such as hyper-pigment, rosacea, skin cancer, etc. (Alam, 2010). The degree of the melanin and haemoglobin can be gained by apply the abortion spectra of the melanin and the haemoglobin to the spectral reflectance of the skin, respectively. The spectral reflectance of the skin can be gained by using TSR or SP. Based on the degree of the melanin and haemoglobin, the skin diseases and the treatment can be more accurately monitored (Takiwaki, *et al*., 2002).

The surface scattering is an important optical character of the skin, which relates to the texture of the skin. The scattering information can be used to model the skin surface structure and skin texture reproduction (Tsumura, *et al*., 2003). For computer graphics, it is an important information for the rendering skin on a 3D image. The bidirectional reflectance distribution function (BRDF) is a mathematic function that defined the property of the light reflected from a surface, which has been frequently used to represent the surface scattering of the skin (Jiang and Kaplan, 2008). The instrument which provides multi-angle illuminations and measurement can be used to determine the values of the parameters in the BRDF. However, it is difficult to use this kind of instrument to measure the skin in vivo, as the skin has non-flat geometry and pressure. So, image-based techniques have been proposed to obtain the scattering information of the skin in previous research work. These methods make use of a mathematical method to predict the texture of the skin through the digital images (Marschner, *et al*., 1999; Tsumura, *et al.*, 2001).

The translucency is a character of the skin, which is found related to the health and the youthfulness of the appearance (Jiang and Kaplan, 2008). The skin translucency is related to the light scattering inside the skin, which is difficult to measure in vivo. Image-based methods were proposed to determine the skin translucency (Jiang and Kaplan, 2008). One proposed method used a digital camera to determine the light refracted out from the skin through the image processing technique. Another image-based method was proposed to change the translucency of the skin (Tsumura, *et al.*, 2008). The images of silicon samples with different translucencies were used to build the mathematical model between the physical translucency and the digital images. Then this mathematical model can be used to change the translucency of the facial image (Tsumura, *et al.*, 2008).

It can be learnt from the above that the melanin and haemoglobin can be gained from the spectral reflectance of the skin. For the surface scattering and the translucency, the digital images are needed.

2.2.2 Skin colour databases

SOCS database

Skin colour databases can provide a reference for evaluating the skin colour reproduction. For investigating the influence of the skin colour to the impression, it can be used to determine the experimental colour range. SOCS (Standard Object Colour Spectra database) is a published colour spectra database containing more than 50,000 items of existing objects, including natural flower, leaves and skin colour. The skin colour data in SOCS were from 6 individual groups, including SHISEIDO, KAO, OOKA, KAWASAKI, OULU and SUN (ISO, 2003). In these 6 groups, only the latter three groups published their system set-up for the skin colour measurement. The SHISEIDO group only contained skin colour of female subjects with makeup and without makeup. The skin colours are all measured by the contacted type of devices. This data group contained skin colour from Thailand, Malay, China and Japan, but most of the skin colour from Japanese. KAO only contained skin colour of Japanese female. OOKA groups used both contact and remote measuring device to collect 63 subjects' skin colour. Most of the subjects were Japanese males (ISO, 2003).

SUN database

SUN skin colour database included the spectral reflectance of 34 subjects' facial skin of (11 Pacific-Asians, 8 Caucasians, 7 Blacks, 6 sub continental-Asians and 2 Hispanics). Ten facial locations were randomly selected two measured at each subject's face. The facial images were captured through a

Sony DKC-ST5 digital photo camera. The skin colour was measured by using a TSR, *SpectraScan 704 (PR-704)* (Sun and Fairchild, 2002). The facial images were also captured during the data collection. But not included in the published database.

OULU database

OULU skin colour database was collected by the University of OULU. The aim of developing this database was to create a database that can be used for a colour related study on human faces (Soriano *et al*., 2000). Similar to the SUN database, OULU skin colour database contained the facial images, which captured by digital camera Sony DXC-755P, and spectral reflectance of skin colour, which measured by a SP (Minolta CM2002). The facial images were captured under four different light sources, including D65, A, TL84 and incandescent. One hundred and eleven subjects, including 100 Europeans, 5 Asiatic and 6 African, were measured and included in this database (Soriano *et al*., 2000). The facial images were not included in the published database.

KAWASAKI database

KAWASAKI skin colour database contains skin colour of 83 subjects, included 80 Japanese, 1 Chinese, 1 Canadian and 1 Finnish. Three facial locations (forehead, zygomatic region and cheek) were measured by using a TSR and a SP (Tajima *et al*., 1998; ISO, 2003). The measuring areas were marked before the measurement to ensure the same location was measured (Tajima *et al*., 1998). Facial images were also captured during the data collection, but they were not available to the public.

Xiao *et al*. database

Apart from SOCS, Xiao *et al*. (2012) built a Chinese skin colour database included 202 Chinese subjects (65 females and 137 males). A SP, Minolta

CM-2600d, was used to measure 6 locations, including the forehead, the tip of the nose, cheek, earlobe, chin, back of the hand, palm, outer forearm and inner forearm (Xiao *et al*., 2012).

Xiao *et al*. (2017) published another skin colour database which contained the skin colour from four ethnic groups (Caucasian (187), Chinese (202), Kurdish (145) and Tai (426)). The data of Chinese is the same as the one which was published in 2012. Two SPs were used to measure the skin colour of two facial locations and two body locations of each subject. The measured skin colours were transformed into CIELAB (D65/2) for analysing the distribution of the skin colour of different ethnic groups in CIE a^*b^* plane and the difference of the L^* of different ethnic groups. They found that the skin distribution of the Caucasian was larger than that of the Chinese in a^*b^* plane. The range of the hue angle and chroma of the Caucasian were larger than the Chinese. The different ethnic groups had the significant difference in b^* value (yellowness) but limited variation in a^* value (redness). They also found that different locations had the difference in a^* value (redness). The body locations were lower in redness, compared with that of the facial locations (Xiao *et al*., 2017).

Even though these databases included the large scale of skin colour data, there are many limitations. Firstly, limited numbers of ethnic groups were included in these databases. Most of the data in these databases were Chinese and Japanese. Secondly, the instruments used to measure skin colour were not unified. Some of the databases used the TSR, the others used the SP. The difference between the results from different types of instruments was uncertain, so it is difficult to use these data together. Even though the same type of instrument was used to measure skin colour, different operations or different measuring conditions can cause the difference in measurement results. Thirdly, limited body locations and a few male subjects were measured in these databases, except Xiao *et al*. databases. Only the skin colours from the facial locations were included in SOCS. Fourthly, these

databases only published the spectral reflectance measured by these two instruments. The facial images were not included.

2.2.3 Skin colour and perceived facial impression attributes

2.2.3.1 Previous studies of the skin colour and perceived facial impression attributes

The previous works investigated the impact of the facial skin colour and the distribution on the *Attractiveness*, *Health* and *Youth*. Some of the studies used the computer generated facial images (CGFI) to investigate the impact of the colour and the distribution of the skin on these attributes. In these study, the CGFI with were used (Fink *et al*., 2006; Fink *et al*., 2008). They found the female CGFI with even or homogeneous skin were appeared younger, healthier, and more attractive (Fink *et al*., 2001; Fink *et al*., 2008). The female CGFI with reddish skin colour were appeared healthier and more attractive (Fink *et al*., 2001).

The other studies used the real human facial images to investigate the impact of the skin colour or texture on the *Attractiveness*, *Health* and *Youth*. The digital camera was used to capture facial images. Then the skin colour and the skin surface topography of the face in the image were altered and used in the experiment. In these study, some of the researchers used the morphed image (Stephen *et al*., 2009a; Stephen *et al*., 2012). The morphed image is an image which was gained via averaging several facial images. The advantage of the morphed image is that the characters of the facial features of all the images are included. So, a morphed image can be used to represent a group of images. The disadvantage of the morphed facial image is that the skin texture is lost after the morphing process. Other researchers used the captured images without the morphing procedure (Stephen *et al*., 2009b; Stephen *et al*., 2011; Samson *et al*., 2010). Some of these researchers used

the polynomial model to correct the colour of the images before the images were used in the experiment (Stephen *et al*., 2009a; Stephen *et al*., 2009b; Stephen, Coetzee and Perrett, 2011). For the examined skin colour, CIELAB coordinates were used to describe the skin colour in the past study. Some of the researchers claimed that the variation of CIE b* value was related to the variation of the carotenoid in the skin (Edwards and Duntley, 1939; Stamatas *et al*., 2004; Stephen *et al*., 2009b Stephen *et al*., 2011; Lefevre and Perrett, 2015). Some of the studies reported that the yellowness of skin was related to the consuming of the fruit and vegetable (Grimm *et al*., 2012; Middaugh *et al*., 2012).

Some researchers studied the health skin colour of the African and Caucasian via adjusting the skin colour of the displayed images. Fifty African images and fifty-one Caucasian images were scaled by about thirty African and about thirty Caucasian observers via adjusting the skin colour of the displayed images in terms of lightness and yellow/redness. The researchers found that the African and Caucasian images with higher yellowness in skin colour appeared healthier (Stephen *et al*., 2011; Stephen *et al*., 2009b; Stephen *et al*., 2009a). Whitehead *et al*. (2012) carried out a similar experiment by using about 30 Caucasian images which observed by 16 Caucasian observers. Researchers also found that the Caucasian with healthy look skin colour also perceived attractive (Stephen *et al*., 2012).

The preferred skin colours were also studied. Some of the researchers found that the skin colour of the Caucasian female with higher chroma value and yellower in hue was preferred than the real skin colour (Sander *et al*., 1959; Bartleson, 1960; Zeng and Luo, 2009). Some of the researchers found that the Oriental females preferred the whiter skin colour (Ashikari *et al*., 2005). The previous study also found that the reddish skin colour appeared whiter than the one with the yellowish skin colour at the high lightness region. Also, the skin colour with the lower chroma appeared whiter than the one with high chroma (Yoshikawa *et al*., 2001). The Yano and Hashimoto (1997) also found

that the preferred skin colour of the Japanese female was higher in saturation and reddish in hue compared to the actual skin colour of the Japanese female. Zeng and Luo (2009) found that the preferred skin colour of Chinese was slightly more reddish than the actual skin colour. Some researchers also found that the preferred skin colour of Chinese was redder than the preferred skin colour of the Japanese (Fan *et al*. 2011). Note that the Fan *et al*. used the printed images in the experiment.

In conclusion, the preferred skin colour of Caucasian women and Japanese women were different. Table 2.2.1 lists the finding of the preferred skin colour of the past studies which rated by the Oriental observers.

Table 2.2.1 Skin colour preference for previous studies (female (F) and male (M); Japanese (JAP) and Chinese (CHN)).

	Stimuli Image	**Observer**	**Preferred skin colour**
Zeng and Luo (2011)	CHN-F (D50)	CHN-F&M	a*b*(20.7, 24.4)
Yano and Hashimoto (1997)	JAP-F (D65)	JAP-F	u'v' (0.2425, 0.4895)
Fan *et al*. (2011)	JAP-F (D50)	JAP-F&M	L*a*b*(74.6, 16.2, 15.2)
Fan *et al*. (2011)	JAP-F (D50)	CHN-F&M	L*a*b*(74.3, 17.6, 12.9)

Apart from the facial impression attributes that were studied, some skin appearance attributes were also investigated in the previous studies. Table 2.2.2 lists the skin appearance attributes that studied.

Table 2.2.2 The skin appearance attributes that investigated in the past studies.

	Attributes
Jiang and Kaplan (2011)	Moisturization
Jiang and Kaplan (2008)	Translucency
Tsumura *et al*. (2008)	Translucency
Wang et al. (2016)*	细腻 (delicate), 水嫩 (fresh), 白皙 (fair),无瑕疵 (flawless), 均匀(uniform),有光泽 (glossy),红润 (ruddy),光滑 (smooth),喜欢 (fond),健康 (healthy)

*the attributes in mandarin were used in their experiments.

2.2.3.2 Other facial impression attributes related to the other facial feature investigated in the previous studies

The impact of the facial features on the facial personality and facial impression was widely studied by psychologist and sociologist. For the facial personality, the Big Five model is a classic model which was used to describe personality (Goldberg, 1990). This model included five factors which are *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism*. Researchers found that these five factors were difficult to directly use in the psychophysical experiment, as these terms were hard to understand by the participants (John and Srivastava, 1999; Pervin and John, 1999). So, the list of adjectives, which were used to substitute the five factors, was developed for experimental use (John and Srivastava, 1999). The *Sociable*, *Cooperative*, *Easygoing*, *Relaxed*, *Imaginative* and *active* were the adjectives that were used to substitute these five factors (McCrae and Costa, 1999).

Many previous studies investigated the relationship between the facial impression and the facial features, e.g. eyes, lips and face. They used the real human facial images to carry out the investigations. Some of them used the facial images with their facial features measured. The others used the images with the facial feature modified by computer software. For the eyes, a previous study used computer modified photographs found that the size of the eyes was found that it related to the *Babyfaceness*, *Attractiveness* and *Femininity* (Paunonen *et al*., 1999). Two previous studies used the original facial photographs found that the larger size of the eyes was perceived as physically weaker, more honest, warmer, more dependent, and more agreeable (Berry and McArthur, 1985; Cunningham, 1986). A previous study used digital images with lip colour altered by computer software to study the impact of the lip colour to the impressions. They found that the colour of the lips related to the *Attractiveness*, and *Feminine* (Stephen and McKeegan, 2010). The female who has redder lips appeared more feminine and more attractive. The female who has bluish lip appeared less attractive and masculine (Stephen and

McKeegan, 2010). The morphed digital images that with the lip altered by the software were used to study the thickness of the lips of the females. They found that the female with thicker lips appeared younger (Gunn *et al*., 2009). For the face, the shape of the face was found that it related to the *Trustworthiness*, *Dominance* and Attractiveness via digital images that captured at the different angle. The ratio of the width and height (fWHR) of the male's face was found to have impact on the *Trustworthiness*, *Aggressiveness*, and *Prejudice* (Carré and McCormick, 2008; Carré *et al*., 2009; Stirrat and Perrett, 2010; Wong *et al*., 2012; Stirrat and Perrett, 2012; Hehman *et al*., 2013b; Hehman *et al*., 2013a). They found that the male with higher fWHR appeared more aggressive, less trustworthy and more prejudiced. The facial symmetry was also found to have an impact on the *Attractiveness* and *Health*. The studies that carried out by using the computer modified digital images found that the symmetry face appeared more attractive than the asymmetry face (Jones *et al*., 2004; Rhodes *et al*., 2007). The symmetry face also appeared healthier than the asymmetry face (Jones *et al*., 2001; Rhodes *et al*., 2001; Rhodes *et al*., 2007).

The attributes mentioned above were important attributes to describe facial impression. The previous work found that these attributes were close related the some of the facial features. But a limited number of studies have investigated the relationship between these attributes and skin colour.

2.3 Whiteness and blackness scale

The whiteness and blackness are commercially important terms to evaluate the quality of the product in various industries, such as teeth, textile and paint. The whiteness and blackness of the skin colour are also frequently used to describe the skin colour in the cosmetic area and they are considered as an important element in Asian beauty (Li *et al*., 2008a). Many formulae that

developed to calculate the whiteness value of the textile, paper and tooth, and to calculate the blackness of the ink were proposed in the previous study. But these formulas only can be used to examine the colour with a limited range of chroma and hue. These formulae are not suitable to be further developed to be used to calculate the skin whiteness.

The Individual Typologic Angle (ITA°) was proposed to determine the melanisation of the skin. The ITA° classify the skin colour in terms of the "lightness" of the skin (Chardon *et al*., 1991; Del Bino, Sandra and Bernerd, 2013; Del Bino *et al*., 2006). The term "lightness" that was used in these studies are not the lightness in CIELAB system. The meaning of the "lightness" is similar to the skin whiteness. The ITA° was developed based on the skin colour distribution in CIELAB L^*b^* plane and the Fitzpatrick skin prototype Type I to VI (see Section 2.2.1). So, the ITA° can be used to determine the relationship between original skin colour and sun tanned skin colour (Chardon *et al*., 1991). The Equation 2.10 is the formula that used to calculate ITA° .

$$\mathrm{ITA°} = \mathrm{ATAN}\frac{(L^* - 50)}{b^*} \times \frac{180}{3.14159} \qquad \textbf{Equation 2.10}$$

where L^* and b^* are the CIELAB L^* and b^* coordinates of the skin colour.

Figure 2.2.3.1 shows the ITA° classification in CIELAB L^*b^* plane. The red arrows that were drawn in this figure indicate the increase of the skin "lightness".

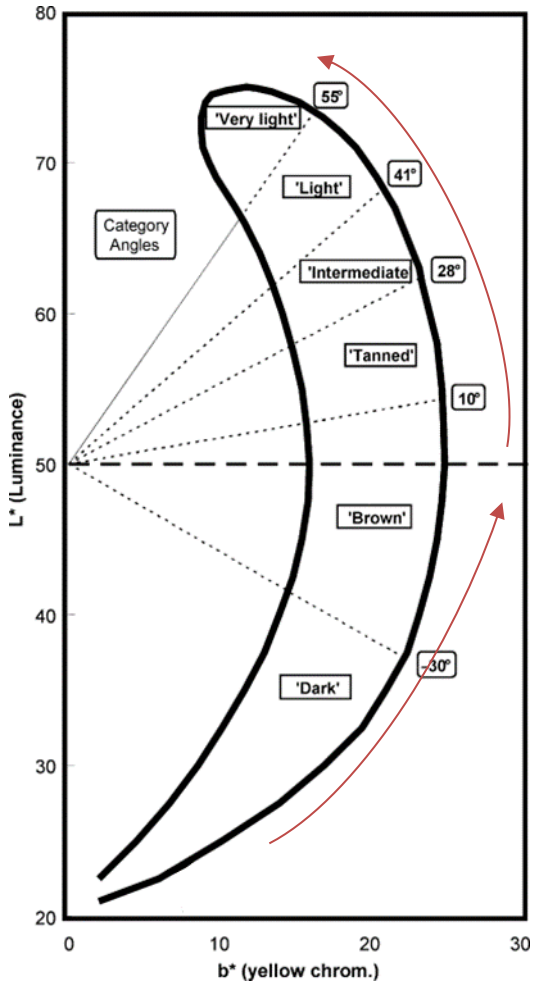


Figure 2.2.3.1 the Individual Typology Angle (ITAº) for skin colour classification (Del Bino *et al*., 2006).

Figure 2.2.3.1 shows that skin colours are classified into six different categories in terms of "lightness" and the skin colour gamut in the L^*b^* is similar to the shape of a banana. The skin colour with higher ITA° appears lighter. The skin colour with smaller ITA° appears darker. In other words, for the skin colours with the L^* value above 50, the skin colours with the higher L^* and smaller b^*values appear lighter. In contrast, the skin colour with smaller $L^*$and higher b^*value appears darker. For the skin colour with the L^* value less than 50, the skin colours with larger L^* and b^* values appear lighter. The skin colour with smaller L^* and b^*values appears darker. The rise of the skin "lightness" can also be expressed by the red arrows in Figure 2.2.3.2.

These red arrows also appeared similar to the meaning of two term-pairs (e.g. cleaner-dirtier and whiter-deeper) that were used to describe the colour variation in the dyeing and paint industries (Berns, 2000, p33). Figure 2.2.3.2 (a) shows the direction of the colour variation that these terms used to describe in CIELAB $L^*C^*_{ab}$ plane. The term-pairs that were used in the different industries were marked in different colour. The term-pairs that were marked in red, green and blue were those used in the dyeing and paint industries (Berns, 2000, p33; Kuehni and Schwarz, 2008, p24).

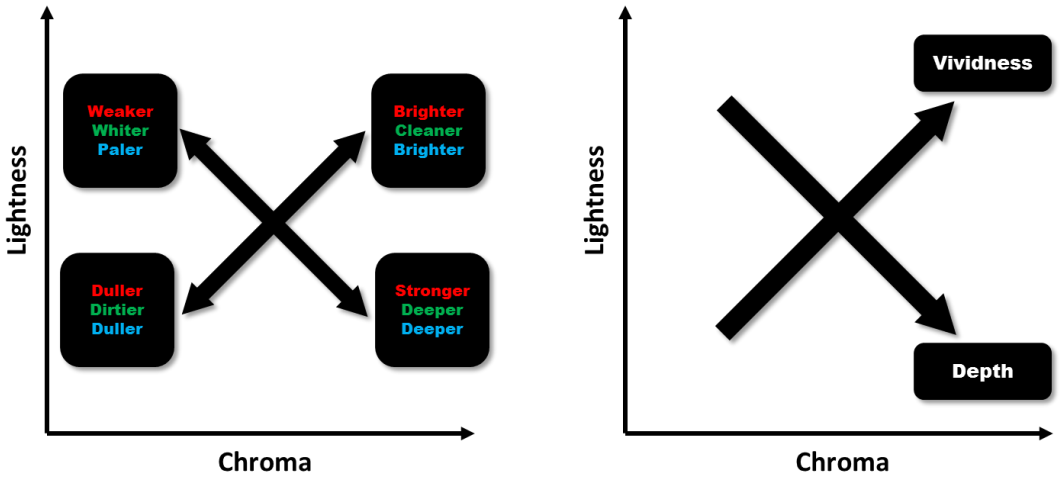


Figure 2.2.3.2 The terms used to describe colour variation in CIELAB $\mathbf{L^*C^*_{ab}}$plane, (a) the attributes used in dyeing and paint industry before Berns (2014): terms in red from dye industry (Berns, 2000,p24); terms in green from paint industry (Berns, 2000, p24); terms in blue from Kuehin and Schwarz (2008,p260 -261). (b) The terms proposed by Berns (2014) which described the colour change in $\mathbf{L^*C^*_{ab}}$ plane (Berns, 2014).

However, the term-pairs in Figure 2.2.3.2, were used to describe colour perception rather than the colour specification, such as CIELAB, of a colour (Berns, 2000,p23-25).

Berns extended this terminology into CIELAB colour space and named *chroma*, *Vividness*, *Depth* and *Clarity* (2014). Table 2.3.1 lists the formulae that transformed CIELAB coordinates into these four scales. The *Vividness*

and *Depth* are used to describe the colour variation in CIELAB $L^*C^*_{ab}$ plane, as illustrated in Figure 2.2.3.2 (b).

Table 2.3.1 The Berns proposed new terminologies (Berns, 2014).

Term	Formula
Vividness	$V^*_{ab} = \sqrt{(L^*)^2 + (C^*_{ab})^2}$
Depth	$D^*_{ab} = \sqrt{(100 - L^*)^2 + (C^*_{ab})^2}$
Chroma	$C^*_{ab} = \sqrt{(V^*_{ab})^2 - (L^*)^2} = \sqrt{(D^*_{ab})^2 - (100 - L^*)^2}$
Clarity	$T^*_{ab} = \sqrt{(L^*_{background} - L^*)^2 + (a^*_{background} - a^*)^2 + (b^*_{background} - b^*)^2}$

As shown in the Figure 2.2.3.2 (b), the vividness of the colour increased when the lightness value and the chroma value of the colour are increased. The depth of the colour increase when the lightness value of the colour is decreased and the chroma value is increased.

Two sets of models that included the formulea for blackness and whiteness are shown below. These two models were developed based on the visual experimental results.

Adams (2010) proposed the blackness (s^+), whiteness (w^+) and chromaticness (c^+) scales to describe NCS colour attributes. The L^* and C^*_{ab} were used to calculated the s^+ and w^+. Equation 2.11 are the formulae that used to calculate the s^+, w^+ and c^+.

$$w^+ = L^* - \left(\frac{C^*_{ab}}{C^*_{ab,pure}}\right)L^*_{pure}$$

$$c^+ = 100\left(\frac{C^*_{ab}}{C^*_{ab,pure}}\right)$$

Equation 2.11

$$s^+ = (100 - L^*) - (\frac{C^*_{ab}}{C^*_{ab,pure}})(100 - L^*_{pure})$$

where $L^*_{pure}$and $C^*_{ab,pure}$ are the lightness and chroma for the lightest colour with the least amount of white for the hue, that is, the colour with the greatest chromaticness (see Figure 2.1.8.2). The sum of w^+, c^+ and s^+ colour attributes is 100.

Cho *et al*. proposed the Ellipsoid-based models for describing saturation, vividness, blackness and whiteness (2017). The model is shown in Equation 2.12.

$$\text{Visual perception} = k_M + k_L\sqrt{(L^* - L^*_o)^2 + k_A(a^* - a^*_o)^2 + k_B(b^* - b^*_o)^2} \quad \textbf{Equation 2.12}$$

where the visual perception was the visual perception of the saturation, vividness, blackness or whiteness. k_L , k_A , k_B , k_M , L^*_o , a^*_o and b^*_o are the constant parameters. The parameters of each models are listed in Table 2.3.2.

Table 2.3.2 The parameters of the saturation, vividness, blackness and whiteness model (Cho *et al*., 2017).

	k_L	k_A	k_B	k_M	L^*_o	a^*_o	b^*_o
Saturation	-1.62	0.05	0.87	0.64	77	2	12
Vividness	-1.81	0.07	0.76	0.38	61	2	16
Blackness	2.35	-0.04	1.47	0.76	0	0	17
Whiteness	1.28	-0.03	0.18	0.22	100	4	-4

2.4 Device characterisation

Reproducing the colour on the monitor consistently and accurately is important for conducting the psychophysical experiment using a display. Figure 2.2.3.1 shows the general procedure of reproducing colour from camera to the monitor.

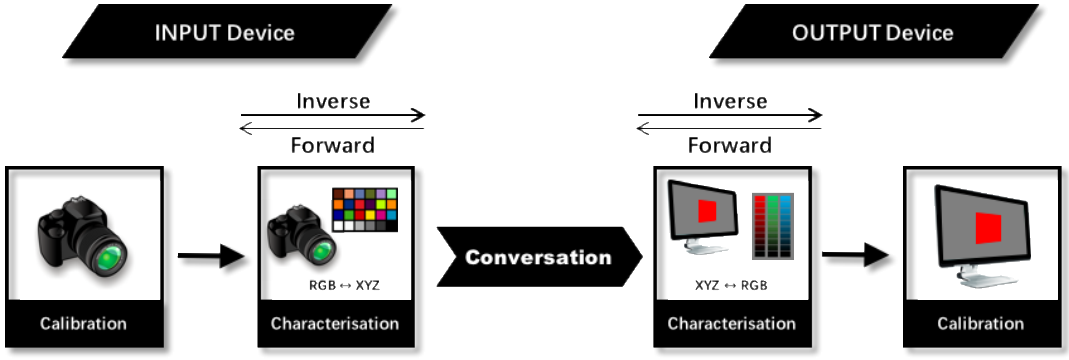


Figure 2.2.3.1 Colour reproduction procedure.

Figure 2.2.3.1 shows two sections in the colour reproduction. They are calibration and characterisation of an input device (such as the camera) or output device (such as the monitor). The calibration procedure is to enable the device to give consistent responses (Sharma and Bala, 2002). This procedure can be achieved by adjusting the setting of the device according to the requirement of the experiment (Sharma and Bala, 2002). Under these settings, the responses of the device should remain consistent.

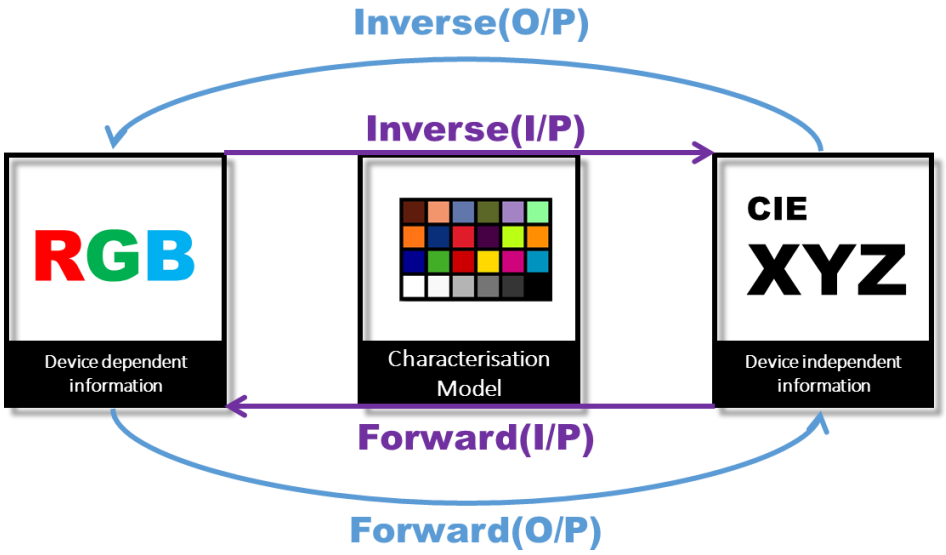


Figure 2.2.3.2 The forward and inverse characterization model: the input device (I/P) marked in purple; the output device (O/P) marked in blue.

The characterisation procedure is to establish a relationship between the device dependent colour signals (such as RGB) and the device independent colour information, such as CIEXYZ (Sharma and Bala, 2002). Two directions of characterisation included in the characterising procedure. One is forward, another is inverse, as shown in Figure 2.2.3.2. For the input device, the forward characterisation is to estimate the response of the device with a known device-independent input, such as CIEXYZ, through the characterization model. The inverse characterization of the input device is to map the device dependent colour information, such as RGB, into the corresponding device-independent colour information, such as CIEXYZ, through the characterization model. For the output device, the forward characterisation is to estimate the RGB signals via the CIEXYZ values. The inverse characterisation of the output device is to map the CIEXYZ values to the RGB signal of the device.

The characterisation models can be classified into 3 categories, the physical models, numerical models and look-up tables (Green, 2002). The physical model is the model that is built based on the physical property of the device. The numerical model is the model expressed as a set of coefficients. The

coefficients are usually generated by a set of the known sample. The look-up table is the method that use a table of the coordinates of the colour space of the device and the related coordinates in the CIELAB colour space to transform the colour.

In this section, a numerical characterisation models for the digital camera and a physical model for the LCD monitor were introduced.

2.4.1 Digital colour camera characterisation

The digital camera is a widely-used imaging device for capturing images. The structure of the camera is similar to the human visual system (Sharma and Bala, 2002). The imaging sensor built inside the camera was used to detect colour information of the capturing target (Sharma and Bala, 2002). Through the imaging sensor, the device dependent RGB information can be recorded. But these device dependent RGB information cannot be directly used to reproduce the colour on the other media. The related device independent colour information is needed. Digital camera characterisation is to build up a relationship between the camera dependent RGB and device independent CIE XYZ of the target colour. So, the colour of the original target can be reproduced.

The regression model is a numerical model which builds up a relationship between the camera dependent RGB and CIE XYZ with a set of coefficients. Two kinds of transformation were commonly used in this model, including the polynomial regression and linear regression (Hunt and Pointer, 2011, p244-249). The polynomial regression was found has higher accuracy than the linear regression (Hong *et al*., 2001; Pointer *et al*., 2001; Cheung *et al*., 2004). The colour mapping that of the polynomial model from RGB to XYZ can be represented by Equation 2.13.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \textbf{Equation 2.13}$$

where M is the transformation matrix that mapping the RGB values to the related CIE XYZ values.

The size of the matrix M was depended on the polynomial terms that were selected to use, which can be 3×3, 3×5 or 3×11. Table 2.4.1 and Table 2.4.2 list the polynomial models with different terms that studied by Hong, Luo and Rhodes (2001) and by Cheung *et al*.(2004).

Table 2.4.1 The polynomial models that were studied by Hong, Luo and Rhodes (2001).

Number of terms	Terms
3	r, g, b
5	$r, g, b, rgb, 1$
6	r, g, b, rg, rb, gb
8	$r, g, b, rg, rb, gb, rgb, 1$
9	$r, g, b, rg, rb, gb, r^2, g^2, b^2$
11	$r, g, b, rg, rb, gb, r^2, g^2, b^2, rgb, 1$

Table 2.4.2 The polynomial models that were studied by Cheung *et al*. (2004).

Number of terms	Terms
3	r, g, b
4	$r, g, b, 1$
5	$r, g, b, rgb, 1$
10	$r, g, b, rg, rb, gb, r^2, g^2, b^2 1$
20	$r, g, b, rg, rb, gb, r^2,\ g^2, b^2, rgb, gr^2, rb^2, bg^2, rg^2, br^2, gb^2, r^3,\ g^3, b^3, 1$
35	$r, g, b, rg, rb, gb, r^2,\ g^2, b^2, rgb, gr^2, rb^2, bg^2, rg^2, br^2, gb^2, r^3,\ g^3, b^3,$ $gr^3, rb^3, bg^3, rg^3, br^3, gb^3, rgb^2, gbr^2, rbg^2, b^2g^2, r^2b^2, r^2g^2, r^4,\ g^4, b^4,$

The transformation matrix M can be achieved through the least-squares fitting method and the training sets which include RGB and related CIE XYZ values. The least-squares can be represented by Equation 2.14.

$$E = \sum_{i=1}^{n} (XYZ_{T,i} - M \times RGB_{T,i})^2 \qquad \textbf{Equation 2.14}$$

where $XYZ_{T,i}$ and $RGB_{T,i}$ are the CIEXYZ and RGB values of i_{th} training sample. The $RGB_{T,i}$ can be in one of the form like the one in Table 2.4.1 or in Table 2.4.2. M is the transformation matrix. n is number of training samples. The matrix M is the one leading E to the minimum. Then, the M can be achieved by solving Equation 2.14, as shown in Equation 2.15.

$$M = ({RGB_T}^T \times RGB_T)^{-1} \times {RGB_T}^T \times XYZ_T \qquad \textbf{Equation 2.15}$$

where RGB_T and XYZ_T are the RGB and CIE XYZ values of the training samples.

The number of orders and training samples were found can influence the accuracy of the polynomial model. For the number of orders, the accuracy of the model increased alone with the increase of the order number (Hardeberg, 2001). But, the accuracy of the model decreased when the order number exceeds a certain value (Li *et al*., 2003). The influence of the combinations of terms on the accuracy of the polynomial regression model was further studied by Hong *et al*. (2001). They found that the polynomial regression function with terms, RGB and 1, improved the accuracy of the model further (Hong *et al*., 2001).

The influence of the number of the training samples was also investigated in the past studies. The accuracy of the model was found had no distinguishable improvement when the number of training samples more than a certain amount (Cheung *et al*., 2004). For the range of the training samples, the training samples that evenly distributed in the colour space were recommended to use (Hung, 1991). The colour charts, such as Macbeth colour checker chart (24 colours) and Munsell book were recommended to use as training samples (Hunt and Pointer, 2011, p244-249). Some of the studies about spectral reconstruction found that the training sample closed to the target can increase the model accuracy (Shen *et al*., 2008; Xiao *et al*., 2016). In the present study, the different number of terms and different range and number of training sample were tested (see Section 3.2.1.1). The most accurate combination was used to characterise the camera.

2.4.2 Monitor characterisation

Different monitor characterisation models were developed based on the requirement of different types of monitor, such as the cathode ray tube (CRT) monitor, the liquid crystal display (LCD) and the LED monitor (Post and calhound, 1989). In the present study, an LCD monitor was used. The LCD monitor has good agreement with CRT monitor at the most characters,

included spatial uniformity, channel independence and stability (Sharma and Bala, 2002; Hainich and Bimber, 2016). So, the models for characterising CRT monitors can be used to characterise some of the LCD monitor (Sharma and Bala, 2002; Hainich and Bimber, 2016). The gamma offset gain (GOG) model is a classic physical characterization model which simulated the process of the transformation between the digital signal and the emitted light (Berns, 1996b; Berns *et al*., 1993; Thomas *et al*., 2008; Thomas and Hardeberg, 2013; Hainich and Bimber, 2016). Equation 2.16 shows an example of the forward and inverse GOG model of the red channel ($[R]$) (Berns, 1996b; Hainich and Bimber, 2016).

Forward $[XYZ] = (K_{gain_r} \times [R] + K_{Offset_r})^{K_{gamma_r}}$

Inverse $[R] = \dfrac{([XYZ])^{1/K_{gamma_r}} - K_{Offset_r}}{K_{gain_r}}$

Equation 2.16

where K_{gain_r}, K_{Offset_r} and K_{gamma_r} are the coefficient of the gain, offset and gamma, respectively.

The K_{gain_r}, K_{Offset_r} and K_{gamma_r} can be obtained through solving the equations with training samples of each channel which included the R, G or B value and the related XYZ.

In these formulae, the output of the monitor when the pixel values equal to zero were considered as zero (completely black). However, in practice, there still light leak out when the pixel values were equal to zero (Hainich and Bimber, 2016). So, the XYZ values that used for the characterisation needed to remove this leaked light first. These leak out light can be achieved through measuring the monitor when setting the pixel values equal to zero.

2.5 Visual Psychophysics

Psychophysics provides a method which quantifies the relationship between the physical measurements of the stimuli and the perception that achieved from the stimuli (Fairchild, 2005, P36).

2.5.1 Chromatic adaptation

CIE define the chromatic adaptation as "*visual process whereby approximate compensation is made for changes in the colours of stimuli, especially in the case of changes in illuminants*" (CIE, 1995a). If the participant is from a lighting environment which is different to the experimental lighting environment, it is important to allow the human participant to take some time for adaption before carrying out the visual experiment, or it is difficult to retain the colour constancy. The time for the chromatic adaptation is crucial for the visual psychophysical experiment. The investigation of the chromatic adaptation time was carried out by many researchers. Some of the researchers studied the time of adaptation from dark environment to a yellowish illumination (Hunt, 1950; Jameson, Hurvich and Varner, 1979; Fairchild and Reniff, 1995). In all cases, the researcher also found that the participants need about 5 minutes to get fully adapted. The researchers also found that longer adaption time was needed when the luminance is also changing (Hunt, 1950). To make sure the observer adapted to the experimental environment, ten minutes adaptation time is normally set for the visual experiment (Barlow, 1972).

2.5.2 Psychophysical experimental method

The psychophysical experiment can be broken down into five sections, including *Stimulus*, *Task*, *Method*, *Analysis* and *Measure* (Kingdom and Prins, 2010 psychophysics a practical introduction). The *Stimulus* is the medium that is used in the experiment. The computer monitor is a preferred medium to use in the experiment. The *Task* is the task that the participants are asked to do in the experiments. The *Task* can be to score images, or as simple as to judge

the visibility of a pattern on the monitor. The *Method* is related to the process of the trials in the experiment, such as the order in which the images appear on the monitor. The *Analysis* refers to the procedure used to convert the data from the experiment into the measurement. For example, the participants score the images with scores from 1 to 7. These scores cannot directly be used in analysis. These scores need to be converted into proportion, then z-scores for further analysis. The *Measure* is a section which aimed build the mathematical model to connect the physical measurement results and the perceptions results.

Many psychophysical methods are proposed to measure human perceptions. Based on the task, these methods can be classified into two categories: adjustment and judgement. The adjustment kind of task require observer to adjust the stimuli to satisfy a criterion which is set by the experimenter. This kind of task requires the criterion is clear, such as matching the reference colour patch. The judgement kind of task asks the observer to judge the stimuli that are presented by the experimenter (Pelli and Farell, 1995). Here, two judgement based psychophysical methods were introduced. These two methods are the *Yes or No* and the *Paired Comparison*. Many previous studies used the *Categorical Judgement* to investigate the impact of the facial features on the impression attributes (Stephen *et al*., 2012; Samson *et al*., 2010; Fink *et al*., 2008; Fink *et al*., 2012; Fink and Matts, 2008; Fink *et al*., 2001; Fink *et al*., 2006; Etcoff *et al*., 2011). The *Categorical Judgement* method is the method that required the observers score images according to the questions. The number of judgements of using *Categorical Judgement* is smaller than that of *Paired Comparison*. The *Paired Comparison* method has better accuracy (Mantiuk *et al*., 2012).

2.5.2.1 Yes or No method

Yes or No is an alternative forced-choice method. Unlike the forced-choice method, the *Yes or No* method only presents one stimulus during each trial (Prins, 2016). A simple example is shown in Figure 2.5.1.1.

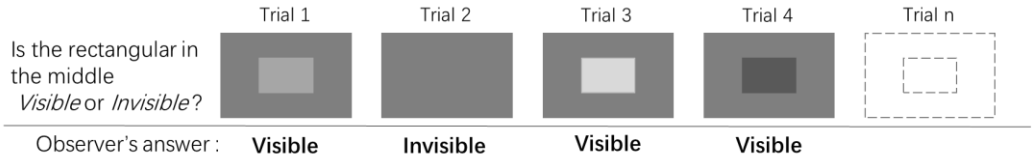


Figure 2.5.1.1 A simple example of *Yes or No* experiment (Prins, 2016).

Figure 2.5.1.1 shows that the *Yes or No* experiment presents only single stimuli. The question in this example was designed as simple *Visible* or *Invisible*. The answer Visible is marked as 1, the invisible is marked as -1. The sum of the response of all observers can be used to determine the visibility of the rectangular with different lightness. For example, an image with altered skin colour was displayed on the monitor with a question, for example, *is the person in the image attractive or unattractive?* Through observing the image, the observer gave the response. The experimental results were transformed into the standard score (z-score) for further investigation. Through the *Yes or No* experiment, the impact of the skin colour to judge each scale can be determined.

2.5.2.2 Paired Comparison method

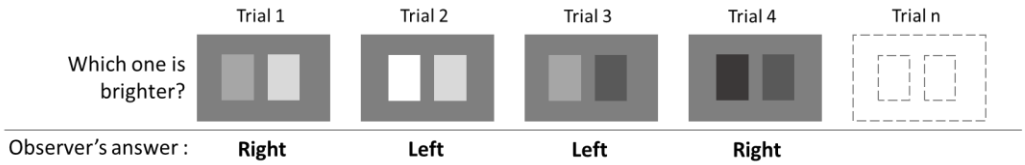


Figure 2.5.2.1 An example of *Paired Comparison* (Prins, 2016).

Paired Comparison method is a simple forced choice scaling method which was used to gain a perceptual scale (Prins, 2016). The *Paired Comparison* method presents two of the stimuli on the monitor. Figure 2.5.2.1 shows an example of *Paired Comparison*. Based on the question, observer chooses the best match stimuli. The chosen stimuli get a response of 1, another one gets a response of 0. This process continued until all the possible combination is examined. The number of pairs can be achieved via Equation 2.17.

$$number\ of\ pairs = \frac{(n^2 - n)}{2} \qquad \textbf{Equation 2.17}$$

where n is the number of stimuli samples.

2.5.2.3 Experimental data conversion

For both method, Thurstone's Case V Model can be used to convert the experimental results into standard score. The standard score is used to transform the measurement results from different observers into the standard scale. *Standard score* (*z-score*) was commonly used to enable the comparison of the results of different observers. The *z-score* shows the difference between the measurement results and the mean of the distribution which gained from the raw data of all observers (Sani and Todman, 2008). The z-score of the paired comparison judgement results can be achieved via the Thurstone's case V Model (Gescheider, 2013). The large *z-score* value indicates the stronger response. The positive and negative sign of the *z-score* indicated the same opinion and opposite opinion to the attributes used in the experiment (Sani and Todman, 2008; Tinsley and Brown, 2000). The procedure to convert the experimental results into z-score is shown below.

The first step is to combine the experimental results of all observers based on the questions. Then, according to Thurstone's Case V Model, the logarithm and proposition were calculated for further converted into z-score (Maydeu-

Olivares, 2004). The equation to calculate the logarithm (LG_i) and proportion (P_i) are shown in Equation 2.17 and Equation 2.18

$$\mathrm{LG}_i = \ln\left(\frac{x_i + 0.5}{N - x_i + 0.5}\right) \qquad \text{Equation 2.17}$$

$$P_i = \frac{x_i}{N} \qquad \text{Equation 2.18}$$

where x_i is the sum of the results of all observer at question i. N is the number of observer.

Then, the inverse of the cumulative standardized normal distribution can be calculated from P_i. The function *normsinv* of Microsoft Excel can be used to accomplish this conversion. The z-score then can be achieved through the linear relationship between the LG and the inverse of the cumulative standardized normal distribution.

2.6 Statistical measures

The statistical methods that used to analyse the experimental results were introduced in this section. The *Microsoft Excel 2011*, *IBM SPSS statistic 23* and *MATLAB R2011a* were used to accomplish the statistical calculation. In these statistical methods, the bubble charts were drawn by using Microsoft Excel. The principal component analysis (PCA), analysis of variance (ANOVA) and correlation coefficient were accomplished by using *IBM SPSS statistic 23*. The mean colour difference from the mean (MCDM), coefficient of determination (R^2) and the standard score (z-score) was calculated by using *MATLAB.*

2.6.1 Mean of colour difference from mean

Mean of colour difference from the mean (MCDM) is used to evaluate the colour variation and stability of the instrument (Berns, 2000, p97; Billmeyer and Alessi, 1981). Typically, the MCDM was used to evaluate the repeatability of the TSR and SP's (Section 3.1.2). The MCDM value is defined as the average colour difference between individual CIELAB values and their average values from a group of the repeated measurements. It can be calculated via Equation 2.19. The larger MCDM value indicates that these measurement results are more spread in CIELAB colour space which means the poorer repeatability.

$$MCDM = \frac{\sum_{i=1}^{n} \Delta E^*_{abi}}{n} \qquad \textbf{Equation 2.19}$$

where

$$\Delta E^*_{abi} = \sqrt{(L^*_i - L^*_{mean})^2 + (a^*_i - a^*_{mean})^2 + (b^*_i - b^*_{mean})^2}$$

$$L^*_{mean} = \frac{\sum_{i=1}^{n} L^*_i}{n}$$

$$a^*_{mean} = \frac{\sum_{i=1}^{n} a^*_i}{n}$$

$$b^*_{mean} = \frac{\sum_{i=1}^{n} b^*_i}{n}$$

n is the number of repeat measurements taken each time. i is the ordinal number of a certain result. L^*_i, a^*_i, b^*_i are the CIELAB values of each measurement.$L^*_{mean}, a^*_{mean}, b^*_{mean}$ are the mean CIELAB value of each group.

2.6.2 Wrong decision

Wrong decision ($WD\%$) is frequently used to evaluate the agreement and accuracy of the technicians at the quality check (McLaren, 1970). The disagreed judgement results are reported as the 'wrong decision', and the

percentage of the disagreed results is reported as the *Wrong decision* (Xin, 2006). The range of the wrong decision's value is from 0% to 100%. The *Wrong decision* can be used to determine the intra-observer and inter-observer variation. For the intra-observer variation, the percentage of the choice that disagreed in the repeat judgment was indicated as the 'wrong decision'. For the inter-observer variation, the percentage of disagreed choice between the observer and majority of the observers. The choice that made by the observer is which different to the majority of the observers at same question was marked as the 'wrong decision'. The smaller wrong decision indicated the better agreement. For the intra-observer variation, the smaller WD% means the observer gave more consistence judgement at this question. For the inter-observer variation, the smaller WD% means that the observers had better agreement at answering this question. Then the agreed results of the observer can be gained, as shown in Equation 2.20.

$$Agreed\% = 100\% - WD\% \qquad \textbf{Equation 2.20}$$

where $Agreed\%$ is the percentage of agreed results.

2.6.3 Principal component analysis

Principal Component Analysis (PCA) is used to investigate the underlying pattern between variables (Abdi and Williams, 2010; Jolliffe, 2002). Through analysing the correlation between variables, the variables can be scored and grouped. One group of variables named a component. The variables within the same component are the variables that strongly correlated. The variables in different component have a weaker correlation. Based on this analysis, the correlation between variables can be determined. Based on this, a variable in a component can be selected out to represent this component. Then, a new set of variables can be generated. This new set of variables has a smaller size than the original variables, but the information of the original variable set can

mostly remain (Cumming and Wooff, 2007; Jolliffe, 2002). The loading matrix were the loading factors of the variable at each component (Jolliffe, 2002). The loading matrix can be plotted based on each component pairs. This visualised the correlation of the variables.

The SPSS can be used to do the PCA analysis. The operation of the SPSS software followed the instructions in book *Introduction to SPSS statistics in psychology* (Howitt and Cramer, 2011).

2.6.4 Analysis of variance (ANOVA)

Analysis of variance (*ANOVA*) is a statistical method that used to investigate the significance of the impact of the independent factor on the experimental results (Sani and Todman, 2008; Tinsley and Brown, 2000). The *p-value* that gained from the *ANOVA* calculation was used to indicate the significance of the correlation between two data sets (Sani and Todman, 2008; Tinsley and Brown, 2000). The *p-value* is no more than 0.05 means that two variables were significantly correlated (Burns and Dobson, 2012; Sani and Todman, 2008; Tinsley and Brown, 2000). This indicates that the variation of the variable has a significant impact on the correlated variable. The SPSS software can be used to calculate the *p-value* (Howitt and Cramer, 2011).

2.6.5 Correlation coefficient

Pearson's product-moment correlation coefficient (*r*) was calculated and used to determine the correlation between attributes (Burns and Dobson, 2012). The formula to calculate r is listed in Equation 2.21.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \textbf{Equation 2.21}$$

where x_i and y_i are the i_{th} samples of two attributes, respectively. $\bar{x}$ and $\bar{y}$ are the average of the samples in each attribute. n is the number of the sample within each attribute.

The absolute value of r close to 1 implies that two attributes have a strong correlation. Contrarily, the absolute value of r close to 0 means two attributes have no strong correlation. The positive r indicated that two attributes have a positive correlation. The positive correlation means one attribute increase in value, another attribute also has increased value. In the contrast, the negative r value indicated that two attributes have a negative correlation. The negative correlation means if one attribute was increased in value, then another attribute was decreased in value. Here, the correlation coefficient was calculated via IBM SPSS. The computer analysis procedure was followed Howitt and Cramer's SPSS statistic instruction book (Howitt and Cramer, 2011).

2.6.6 Coefficient of determination: R^2

The *Coefficient of determination* (r^2) is the square of the correlation coefficient which used to indicate the similarity of the variance of two data sets (Tinsley and Brown, 2000). It is also widely used to determine the distance between the data and their regression line (Tinsley and Brown, 2000). The r^2 can be calculated by Equation 2.22 (Watier *et al*., 2014).

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(f(x_i) - Y_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$ **Equation 2.22**

where Y_i is the y coordinate of the data point. n is the number of the data points. $f(x_i)$ is the trend line function. $\bar{Y}$ is the average of the data points' y coordinates.

r^2 close to 0 indicated that the poor agreement between the data and the trend line. In the contrast, the r^2 value close to 1 means the data and the trend line have an acceptable agreement.

2.6.7 Bubble chart

The bubble chart is used to illustrate three-dimension data in a two-dimensional diagram. In the present study, bubble chart was used to visualise the relationship between the *z-score* of each attribute and two skin colour scales, skin whiteness and hue angle (see Section 5.2.2.2). Figure 2.6.8.1 shows an example of the bubble chart.

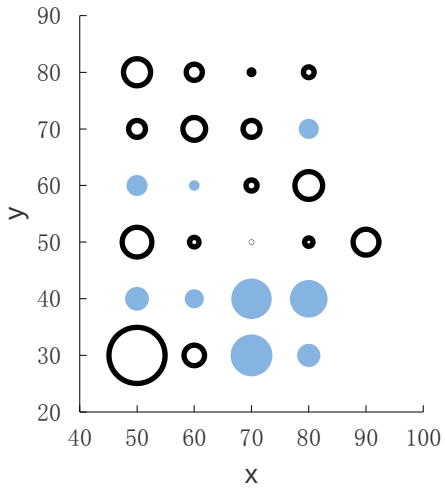


Figure 2.6.8.1 An example of the bubble chart.

As shown in the figure, the white bubble indicated the negative *z-score*, the coloured bubble indicated the positive *z-score*. The size of the bubble indicated the absolute value of the *z-score*. In the bubble chart, the larger coloured bubble was shown that the observer gave a strong response.

2.7 Summary of Chapter 2

This Chapter first reviewed the colorimetry. In this section, the CIE D65 illumination, the di:8° measurement geometry and the measurement instruments (TSR and SP) were related to the present study. The formulae that used to transfer the spectral power distribution to the spectral reflectance, the formulae that transformed the spectral reflectance into CIEXYZ and then into CIELAB were also included. The psychophysical experimental methods, the *Yes or No* and *Paired Comparison* that used in the present study were reviewed in Section 2.5. The statistical measures that used to evaluate the observer's behaviours and to calculate the experimental results for further analysis were included in Section 2.5.

The second section reviewed the skin colour measurement methods and the published skin colour databases. The skin colour measurement methods that included the skin colour charts, TSR and SP, which were also the methods that used to measure skin colour in the present study. And the latter two instruments were used to collect skin colour data to build the skin colour database. The published skin colour databases have four main limitations. Firstly, limited variety of ethnic groups' data were included in these databases. Secondly, the instrument used to measure skin colour were not unified. Thirdly, limited body locations and a number of male subjects were measured in these databases. Fourthly, limited types of data included in these databases. So a new database was needed. The past studies that investigate the impact of the skin colour were also reviewed in this section. These studies were classified

based on the different experimental stimuli, the computer generated facial images and real human images. The limitations of the previous studies can be concluded into two parts. The first one is the limited skin colour range were investigated. The second one is the limited perceived facial impression was examined. So, the present study was carried out to overcome these limitations.

The third section reviewed the whiteness and blackness related indices. The Vividness and Depth formulae were adopted in the present study to develop a skin whiteness and a skin blackness indices.

The fourth section reviewed the camera characterisation method and monitor characterisation method that used in the present study. The accuracy test methods were also reviewed in this section.

Chapter 3 Experimental

In the present study, two experiments were carried out. Experiment 1 aimed to build a skin colour database. Three measurement methods included two instruments and a visual assessment. Experiment 2 aimed to investigate the impact of the different skin colours on the perceived impression of those skin colours. The digital images, which were captured by a digital camera, displayed on a monitor for experimental use. The measurement instruments and the imaging devices used were evaluated prior to the two experiments. The imaging devices were also characterised to capture and display accurate colour on the monitor. This chapter reports the evaluation methods and results and the characterisation models for the instruments and devices used in the present study.

For the skin colour measurement, a light booth and two instruments, a tele-spectroradiometer (TSR) and a spectrophotometer (SP) were used. The quality of the light source is described in Section 3.1.1. For the two instruments, the repeatability of theirs in measuring skin colour at different instrument settings and operation methods is evaluated in Section 3.1.2. Based on the repeatability, the measurement protocols for these two instruments are described. Also, the differences between the measurement results of the two instruments are also reported.

For the facial impression experiments, the imaging system was used to capture facial images and, by image processing, these images were used to generate the stimuli images. The stimuli images were displayed on a monitor for the observer to make judgements. To accurately display the desired skin colour of the image on the monitor, the camera and the monitor were characterised. The accuracy of the camera characterisation models with different settings was evaluated and the most accurate model was used to characterise the device.

3.1 Experimental preparation for Skin Colour Measurement

3.1.1 DigiEye facial image light booth

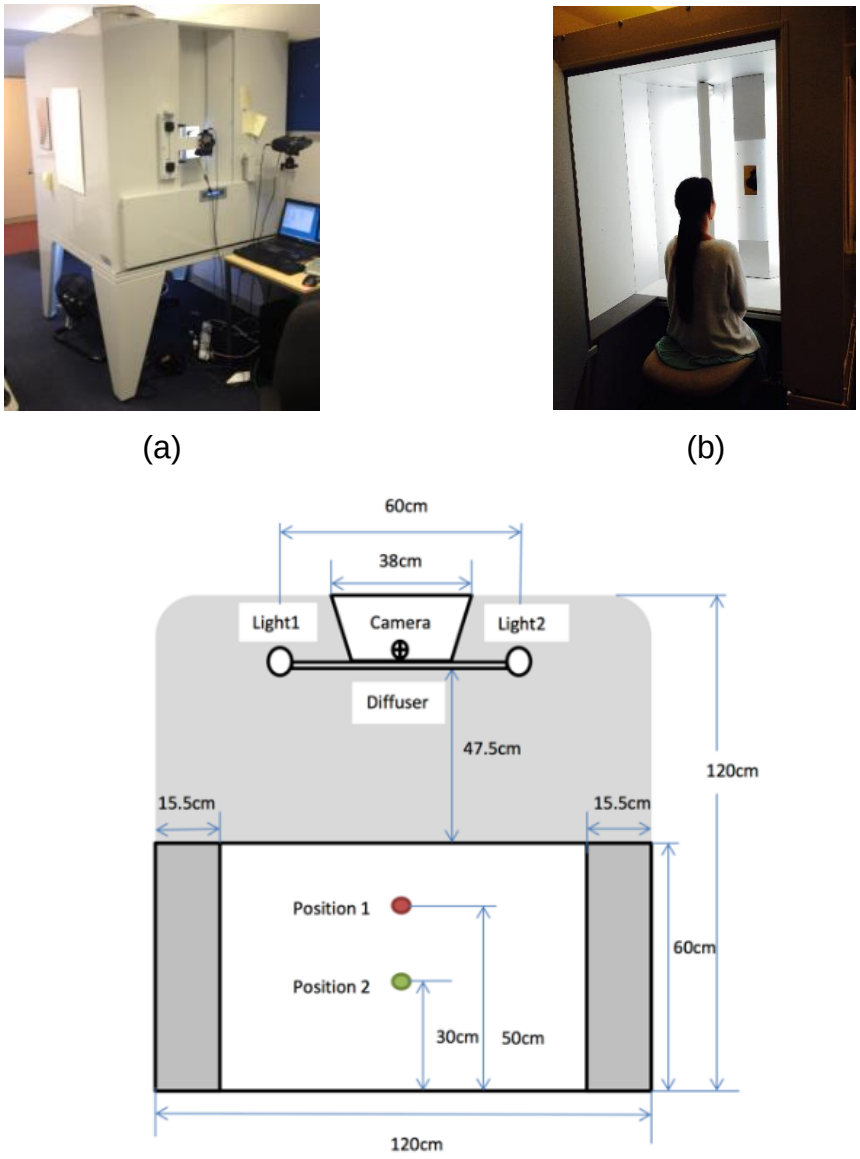


Figure 3.1.1.1 The DigiEye facial image light booth. (a) Outside; (b) Inside; (c) Structure of the light booth and the subject locations.

The *Verivide*® *DigiEye facial image system* included two components, the illumination system (*DigiEye facial image* light booth) and the imaging system (*DigiEye Facial imaging system*) (the imaging system is described in Section 3.2.1). The illumination system is a 120 cm x 120 cm x 200 cm light booth placed on a 100 cm high base, which allowed the participants to sit inside, as shown in Figure 3.1.1.1 (b). This light booth is painted a neutral matt colour inside and a D65 fluorescent simulator is placed inside the booth to provide diffuse illumination. This light booth was also used to provide illumination for the TSR for measurement and to carry out the visual assessments. Behind the illumination, two mechanical arms were used to hold the camera and the TSR. These two arms allowed the two instruments to be interchanged while keeping the same measuring and capturing positions.

Six properties of the light source of the light booth were investigated, including the warm-up time, correlated colour temperature (CCT), luminance, illuminant white point, relative spectral power distribution and colour rendering index. The properties of the light booth were determined through measurement of a white tile. The TSR (see Section 3.1.2.1) used in the present study can provide the CCT, luminance and SPD as part of the instrument output.

The built-in warm-up timer of the light source was set to give a warm-up time of 10 mins. To determine the actual warm-up time of the light source, the luminance of a white tile was measured, using the TSR, 16 times with 2 mins intervals for a total of 30 mins, as shown in Figure 3.1.1.2. This figure shows that the light source does indeed appear to be stable after approximately 10 mins. So, the warm-up time set by the light source can be used to monitor the warm-up of the light source. To ensure the light source was warmed up, the experimenter set warm-up time to be 30 mins.

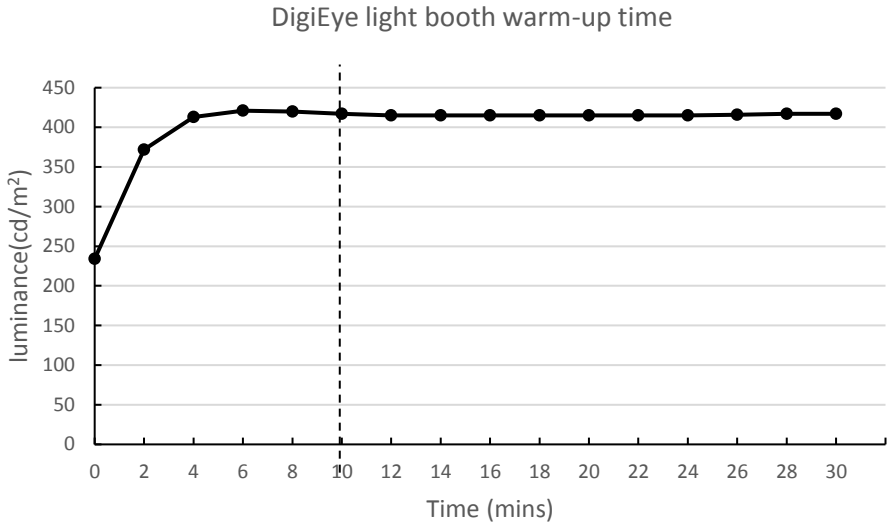


Figure 3.1.1.2 The change in luminance during the DigEye facial image light booth warm-up (30 mins, 2 mins interval).

Apart from the warm-up time, the CCT, the luminance and the white point of the light source were also investigated by measuring the same white tile sequentially five times with two different measurement distances, 57.5 cm (position 1) and 77.5 cm (position 2), as shown in Figure 3.1.1.1 (c). The measurement was carried out after turning the light on for 40 mins. Table 3.1.1 lists the mean CCT and the mean luminance of the five repeat measurements that were taken at the two distances.

Table 3.1.1 The correlated colour temperature and luminance of the light booth.

	CCT.	Luminance (cd/m²)	x	y
Position 1	6726K	628	0.30634	0.32897
Position 2	6771K	416	0.30543	0.32786

Table 3.1.1 shows that the correlated colour temperature at different measurement distances was very similar. The luminance at Position 1 was

50% higher than that at Position 2. Because Position 1 was closer to the light source. The white points at the two positions had similar values of x, y chromaticity. The relative SPD of the illumination in the light booth is shown in Figure 3.1.1.3. The relative SPD shows a good coverage of the visual range of the spectrum.

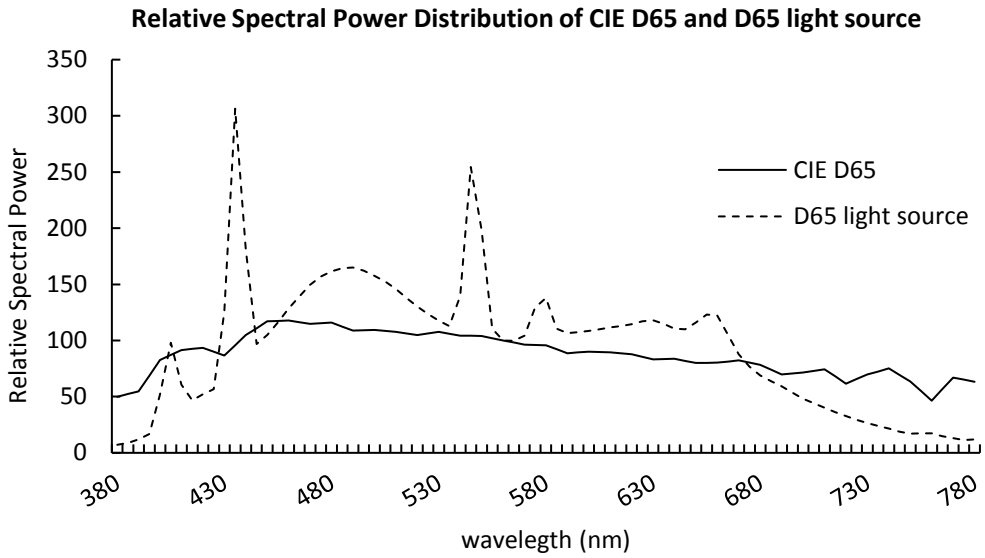


Figure 3.1.1.3 The relative SPD of the light source in DigiEye facial image light booth and the CIE standard illuminant D65.

The CIE colour rendering index (Ra) was used to evaluate the colour quality of the light source. The higher the Ra value the better the colour rendering quality. The ideal value of Ra for the standard illumination is 100. The Ra of the DigiEye light booth was calculated via Equation 2.1 in Section 2.1.1. It was found that the Ra of the source in the light booth was 98. This value indicates that the illumination provided by the DigiEye light booth had an excellent colour rendering ability.

3.1.2 Colour measuring instruments

Different from paper and plastics, the human skin has a multilayer structure, which is sensitive to stretch and pressure, and has an uneven colour on the surface. Different measuring methods and operation routines routine could affect the result. For example, a different measurement field size can affect the measurement results, because of the uneven distribution of colour over the surface of the skin. Different sizes of contact area and lengths of contact time can also affect the measurement results, as human skin is sensitive to pressure. In other words, the measurement field size, contact area size and measurement time can affect the measurement results. There is very little research that has evaluated these uncertainties in the measurement of human skin colour. Therefore, before accumulating skin colour data, the uncertainties associated with the use of a SP and a TSR, when used to measure skin colour, were investigated. Measurements were made of skin colour patches at two locations, the forehead and the cheekbone, on the human face. In this section, the TSR and SP used in the present study are first introduced. The short-term repeatability of the different instruments at different settings are described. Finally, the colour shift between measurements from different locations, between different ethnic groups, and between different instruments are reported.

3.1.2.1 Tele-spectroradiometer

The TSR used in the present study was a *Photo Research SpectraScan PR650* (PR650), as shown in Figure 3.1.2.1. This is a non-contact colour measuring instrument that can measure the skin colour without disturbing the surface of the skin. The measurement geometry, however, is quite different from that of a contact type of instrument, due partly to the unavoidable movement of the participant during the measurement. Especially for

measuring a subject with a dark skin colour, the instrument requires a longer integration time. The size of the measuring area depended on the measurement distance, as the measurement angle of the TSR was fixed (1°). The distance between the target area and instrument was monitored before making a measurement. The TSR was located at the back of the DigiEye facial image light booth and held by a mechanical arm, as shown in Figure 3.1.1.1 (a). The measurement geometry of the TSR was such that it measured a field size that subtended 0°:0°. To achieve this measurement geometry, the TSR was kept still and the participant sat inside the light booth. The height of the seat was adjusted to allow the participant's head to be in the correct position relative to the instrument. Measurement results from the TSR were reported in terms of the SPD with a wavelength range from of 380 nm to 780 nm at intervals of 4 nm.

(a) (b)

Figure 3.1.2.1 Photo of Research SpectraScan PR650. (a) Side view; (b) Back view. (TERAPEAK, 2014).

3.1.2.2 Spectrophotometer

(a) (b)

Figure 3.1.2.2 *KONICA MINOLTA CM-700d* (CM700d) and measurement environment. (a) CM700d; (b) measurement environment.

The SP used in the present study was a *KONICA MINOLTA CM-700d* (CM700d), as shown in Figure 3.1.2.2 (a). The CM700d is a portable contact type instrument with a built-in integrating sphere system. The illumination of the instrument was a built-in flash with a UV cut-off filter. In the present study, a D65 simulator and the measurement geometry of di:8° was used. The CM700d was calibrated before taking any measurements. By following the manufacturer's instructions, a white tile, which was provided by the manufacturer, was used to carry out calibration. The zero calibration was carried out in a dark environment with no reflective items within 1 metre. This instrument provides measurement results in terms of spectral reflectance with a wavelength range of 400 nm to 700 nm at an interval of 10 nm. Four changeable target masks are included in the instrument package. These masks are two with different measuring field sizes and two with different contact area sizes, as shown in Figure 3.1.2.3 (a)-(d). The different aperture sizes, Φ 8 mm and Φ 3 mm, were referred to as MAV and SAV respectively. Each aperture size is coupled with two different target masks: one with a plate in front to reduce the pressure on the measurement surface and the other without the plate, referred to as LP and HP, respectively. Human skin colour

measurement results might vary with different measuring field sizes and/or different applied pressures. Therefore, before the main data accumulation process, the short-term repeatability of using the different masks was investigated.

(a) (b) (c) (d)

Figure 3.1.2.3 Four target masks of the CM700d. (a) SAV-HP; (b) MAV-HP; (c) SAV-LP; (d) MAV-LP.

3.1.2.3 Data processing

As the PR650 and the CM700d provided different formats for the measurement results, it was necessary to transform the data into the same format for further analysis. The measurement results of the PR650 were the SPD which had the wavelength range of 380 nm to 780 nm at 4 nm intervals. The measurement results of the CM700d were the spectral reflectance which had the wavelength range of 400 nm to 700 nm at 10 nm intervals. In the present study, all the measurement data were stored in a fixed format which was the spectral reflectance with the wavelength range of 400 nm to 700 nm at 1 nm intervals. A 5th Order Sprague Interpolation was used to interpolate the measurement results from 4 nm and 10 nm intervals to 1 nm intervals (CIE, 2005).

To transform the SPD, as measured by the PR650, into spectral reflectance, the SPD of a perfect reflecting diffuser, measured at the same location under the same light source (for non-self-luminous sample) using the same instrument, was needed. The perfect reflecting diffuser, however, is not

accessible for practical measurement (see Section 2.1.7). But the SPD of the perfect reflecting diffuser can be gained via a mathematical method. As described in the Section 2.1.7, the spectral reflectance measured using the PR650 also required the SPD of the perfect reflecting diffuser but these data were provided as part of the manufacturer's calibration of the instrument. Thus the SPD measured by the PR650 could be transformed into spectral reflectance via the CM700d calibration data. Equation 3.1 shows the transformation from the measured SPD to the equivalent spectral reflectance.

$$R_{\text{sample}} = \frac{SPD_{sample}}{SPD_{p.d}} \qquad \textbf{Equation 3.1}$$

where $SPD_{p.d} = SPD_{w.}/R_w$, and SPD_w is the SPD of a working standard white tile measured by the PR650. R_w is the spectral reflectance of the same working standard white tile measured by the CM700d. SPD_{sample} is the SPD of the sample. Note that the R_w is normalised against the perfect standard, the perfect reflecting diffuser.

After transforming all of the measurement results into spectral reflectance, CIEXYZ tristimulus values could be calculated, with a subsequent calculation of CIELAB (D65/2) coordinates (Equation 2.2 and Equation 2.5) for further analysis.

3.1.2.4 Evaluation tests design

Two tests were designed to investigate the short-term repeatability and colour shift using both instruments. The first test (Test 1) aimed to determine the short-term repeatability and colour shift of these two instruments through measuring paper-based skin colour patches. Five skin colour patches were measured by both instruments at different settings. For the colour shift, 110 skin colour patches were measured by both instruments at different settings.

The second test (Test 2) aimed to determine the short-term repeatability and colour shift within and between these instruments using different settings while measuring skin colour *in vivo*. In this test, two facial locations, the forehead and the cheekbone, of eleven subjects, including 7 Caucasians, 3 Chinese and 1 South-Asian, were measured. To ensure the two instruments were measuring the same location, the two facial locations were marked before the measurement. The settings of the PR650 and the CM700d are listed in Table 3.1.2.

Table 3.1.2 The test settings of PR650 and CM700d.

Instrument	Aperture size	Contact area size	Contact time
PR650	1. Φ10 mm (measuring distance 575mm) 2. Φ14 mm (measuring distance 775mm)	N/A	N/A
CM700d	1. Φ8 mm(MAV) 2. Φ3 mm(SAV)	1, large contacting area (with plate in front)-LP 2, smaller contacting area (without plate in front)-HP	1, continuous repeatability measurement (CT) 2, consecutive repeatability measurement (CS)

For the PR650, as the measurement angle is 1 degree, different measuring distances can result in different equivalent measuring field sizes. In these tests, two measuring distances were set. These two distances were 57.5 cm (measuring field Φ10 mm) and 77.5 cm (measuring field Φ14 mm), named position 1 (P1) and position 2 (P2), respectively. The measuring distance was the distance from the lens of the TSR to the subject's face.

For the CM700d measurements, three sets of instrument settings were investigated. These settings included two different aperture sizes, two different pressures (different contact area sizes) and two operational methods. For the first two, four different target masks, as described in Section 3.1.2.2, were used to achieve the measurements that differed in measuring field size

and in applied pressure. For the latter setting, all four masks were used to measure each facial location with two operational methods, the continuous repeatability measurement (CT) and consecutive repeatability measurement (CS). For the former, the facial location was measured 5 times continuously without moving the instrument. For the latter, the facial location was also measured 5 times, but the instrument was removed from the subject between measurements.

To assess the agreement between the skin colour measurements, all the measurement results were transformed into CIELAB values (D65/2). The short-term repeatability was investigated by calculating the MCDM value of the five continuous measurements. Section 2.6.1 fully describes the calculation of the MCDM. The colour shifts between the same different settings of the instrument and between the two instruments were investigated through plotting the average measurement results in CIELAB a^*b^* and $L^*C^*_{ab}$ plane and reporting the results in terms of the colour difference in ΔE^*_{ab} units.

3.1.2.5 Results and Discussion

3.1.2.5.1 Test 1 measuring colour patches

Short-term repeatability

The short-term repeatability was evaluated through the measurement results of 5 randomly chosen paper-based skin colour patches from the *Pantone SkinTone™ Guide*. The mean MCDM values of the measurement results from the 5 skin colour patches are listed in Table 3.1.3. The mean MCDM values were 0.065 and 0.085 ΔE^*_{ab}, for measurements made using the CM700d and the PR650 respectively. These values can be used to represent the typical short-term repeatability of these two instruments respectively.

Table 3.1.3 The mean of the MCDM of 5 *Pantone SkinTone™ Guide* patches with 5 repeat measurements. (PR650 and CM700d).

Mean MCDM (ΔE*ab) of	PR650	CM700d
5 *Pantone SkinTone™ Guide* patches	0.09	0.07

Cross instrument comparison

The colour shift between the two instruments was evaluated through measuring all 110 skin colour patches in the *Pantone SkinTone™ Guide*. The mean colour difference between measurements from the two instruments is listed in Table 3.1.4. From the mean colour difference, it can be seen that the difference between the two instruments below 1 ΔE^*_{ab}. The difference between the two instruments can probably be attributed to the difference in measurement geometry. The measurement geometries of the CM700d and the PR650 were di:8° and 0°:0°, respectively.

Table 3.1.4 The mean colour difference between the two instruments when measuring skin colour patches.

Colour difference	Mean(ΔE*ab)	Standard deviation
110 *PANTONE Skintone Guide*® patches	0.88	0.17

3.1.2.5.2 Test 2

Short-term repeatability

The short-term repeatability of CM700d and PR650 with different settings were investigated through the MCDM of 5 repeat measurements of two facial locations. The mean MCDM values of each instrument's settings are listed in Table 3.1.5.

Table 3.1.5 The short-term repeatability of the CM700d and PR650 with different settings at measuring real human facial locations.

	PR650		CM700d					
	Position 1	Position 2	Operational methods		Different pressures		Different aperture sizes	
			CT	CS	LP	HP	MAV	SAV
Mean MCDM(ΔE^*_{ab})	0.51	0.53	0.37	0.63	0.34	0.40	0.35	0.39

As shown in Table 3.1.5, for the TSR, the mean MCDM values for both measurement distances were approximately 0.50 ΔE^*_{ab}. This indicated that the measuring distance (measuring field size) has limited effect on the short-term repeatability of the PR650. For the CM700d, the mean MCDM value of the CT method was almost half that of the CS method. The mean MCDM values using different pressures and different measuring field sizes were similar. This indicated that, as might be expected, the operational methods have an impact on the short-term repeatability of CM700d measurements. The operational method CS was less repeatable than the CT method. Different pressures and measuring field sizes have limited impact on the short-term repeatability. The measurement results of the large aperture size (MAV) and low pressure (LP) target masks have slightly better repeatability. In Table 3.1.5, the mean MCDM values for the LP and HP target masks were 0.34 and 0.40 ΔE^*_{ab}, respectively; the mean MCDM values for the MAV and SAV target masks were 0.35 and 0.39 ΔE^*_{ab}, respectively. This implies that different measuring field sizes and pressures could result in similar short-term repeatability.

Cross instrument comparison and the influence of different instrument settings

The mean colour difference between measurements obtained using different settings of the same instrument and between different settings of the two different instruments are listed in Table 3.1.6. These colour differences were the average of the colour difference between the measurement results using the two instruments at the same facial locations.

Table 3.1.6 Colour difference between different settings and between different instruments.

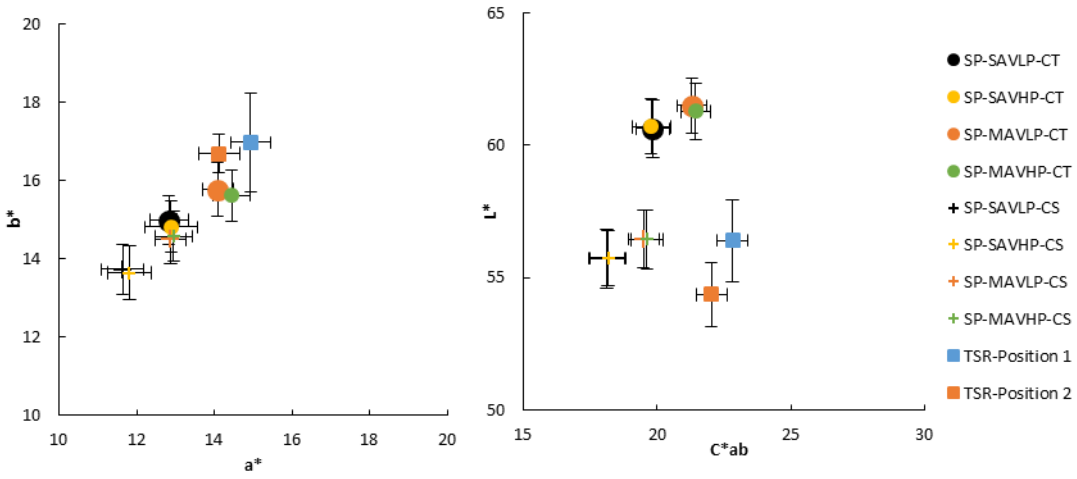
Measurement instrument(s)	Cross-compare pair	Mean Colour Difference (ΔE^*_{ab})
PR560	**Position 1 and Position2**	2.79
CM700d	**LP with different aperture sizes (MAV and SAV)**	2.48
	HP with different aperture sizes (MAV and SAV)	3.04
	MAV with different pressures (HP and LP)	0.88
	SAV with different pressures (HP and LP)	1.82
PR650 and CM700d	**PR650 and CM700d (SAVLP)**	3.49
	PR650 and CM700d (MAVLP)	2.57
	PR650 and CM700d (SAVHP)	4.02
	PR650 and CM700d (MAVHP)	2.69

For the PR650, the mean colour difference between the measurements made at P1 and at P2 was approximately 2.79 ΔE^*_{ab}. For three different setting sets of the CM700d, the mean colour differences between different measurement pressures were 1.82 ΔE^*_{ab} and 0.88 ΔE^*_{ab}, for SAV and MAV respectively. The mean colour differences between two measuring field sizes were 2.48 and 3.04 ΔE^*_{ab}, for LP and HP respectively. These results indicated that the measuring field size had a greater impact on short-term repeatability than

different pressures. These findings further imply that the human skin colour is not evenly distributed and reasonable pressures on the skin do not greatly affect the measurement result.

The mean colour differences between the PR650 (P1) and CM700d (four different masks) are listed in the last four rows of Table 3.1.6. The colour difference between the measurement results of the PR650 (P1) and the CM700d (MAV) were 2.57 and 2.69 ΔE^*_{ab}, for LP and HP respectively. The colour difference between the measurement results of the PR650 (P1) and the CM700d (SAV) were 3.49 and 4.02 ΔE^*_{ab}, for LP and HP respectively. These indicate that the larger aperture size masks resulted in a better agreement with the PR650 measurement at P1, as the measuring field size of the PR650 at P1 (Φ10 mm) was similar to the MAV (Φ 8 mm). From these test results, it can be found that different types of measurement instruments gave different measurement results. Furthermore, this also indicates that the measuring field size can greatly affect the measurement result. The large difference between the two measuring field sizes of the two instruments could contribute to an increase in the colour difference between their measurements. The colour difference between the PR650 and CM700d is similar to the finding of Kawasaki (Tajima *et al*., 1998), where they found the colour difference between TSR and CM700d about 4 ΔE^*_{ab}

The mean CIELAB values of the two instruments with each of a total of 10 different settings, including 8 for the CM700d and 2 for the PR650, were investigated by plotting the results in CIELAB a^*b^* and CIELAB $L^*C^*_{ab}$ planes, as shown in Figure 3.1.2.4.



(a) (b)

Figure 3.1.2.4 Colorimetric shift between different settings of the PR650 and the CM700d: (a) 10 settings plotted in a $\boldsymbol{a^*b^*}$diagram; (b) 10 settings plotted in a $\boldsymbol{L^*C^*_{ab}}$ diagram.

In these figures, each data point is the mean CIELAB value of each of 2 facial locations of 7 subjects that were measured with different instruments and settings. For the PR650, the measurement result at P1 (the shorter measurement distance) had higher lightness and chroma values and appeared redder than the measurement at P2 (the longer measurement distance). This can because of the difference in measuring field sizes. The measuring filed size at P2 is larger than the one at P1, which indicated that the measurement results of P2 including more skin colour information of the location, such as the redness, dark spots, etc. For the CM700d, the colour shift between the two different operational procedures was significant. For the CT method, the measurement results had higher lightness and chroma than that for the CS method, but their hue angles were similar. This can be caused by the different measuring area. For the CS method, the shift of the measuring field cannot be avoided between the measurements of the same location, as the CS method needs to move the instrument away from the skin between two repeat measurements. This can result in the measuring area of the CS is slightly larger than that of the CT. The measurement results for the HP masks

appeared redder than those of the LP masks. This can be because the pressure affected the blood circulation of the measurement area. The mask of the CM700d is a ring shape. The higher pressure lead to the area around the measurement area have worse circulation, which can result in the measurement area appeared redder. The measurement results from the MAV were higher in chroma and lightness. This can be caused by the difference in measuring field size, as the MAV have larger measuring field size, which included more skin colour information. The above can be concluded that different measuring field size results in colour appearance shifts in chroma and lightness. The measurement results from the MAV were higher in chroma and lightness. The measurement results for different measuring pressure had an insignificant difference in the lightness and chroma.

3.2 Facial impression experimental preparation

3.2.1 Facial imaging system

In the present study, a *DigiEye Facial imaging system* was used to capture facial images that could be used in the psychophysical experiment. The *DigiEye facial imaging system* is part of the *Verivide® DigiEye facial image system*. Another component of the system, the light booth, was described and evaluated in Section 3.1.1. The facial imaging component includes a camera and computer-based capturing software, *DigiEye V2.7.1.0*. The camera used in this research was a DSLR camera, a *Nikon D7000*, as shown in Figure 3.2.1.1(b). In this system, the camera is connected to a computer and remotely controlled by the installed software to achieve image capture under controlled conditions. The shutter speed and the aperture could be set in the software and were set to have values of 1/20 sec and f/7.1, respectively. These values were recommended by the *DigiEye facial imaging system*. The

focal length was set to be 50 mm to allow the whole facial area of the participant to be captured.

(a)

(b)

Figure 3.2.1.1 The calibration colour chart and the digital single-lens reflex camera used in the imaging system. (a) DigiTizer chart V3.3; (b) Nikon D7000.

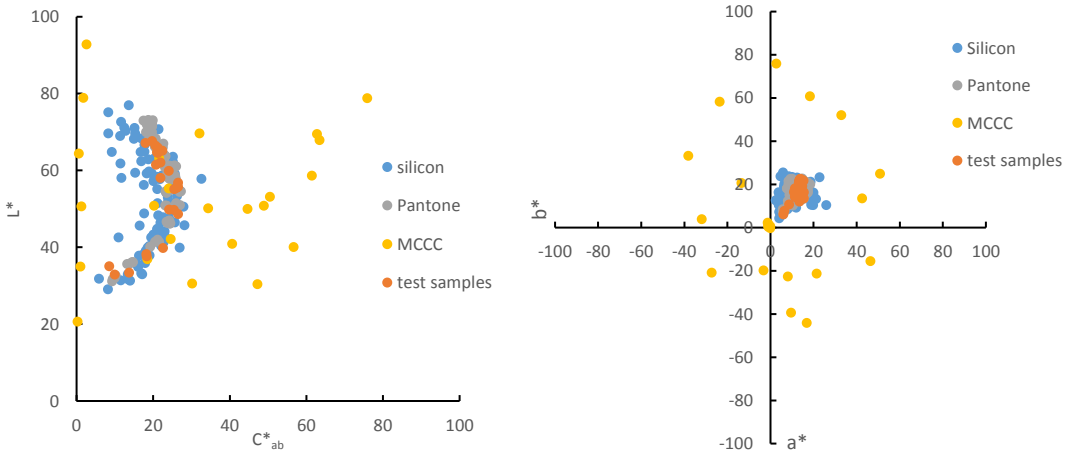
Before capturing, calibration was carried out by imaging a matt white flat board which was of such a size to cover the imaging area, and the DigiTizer colour chart, which was specially designed for the *DigiEye facial image system* and supplied with the system, as shown in Figure 3.2.1.1 (a). The calibration was achieved by following the *DigiEye facial imaging system* instructions. The whiteboard was used to determine the image white balance and to achieve a correction for any non-uniformity of the illumination over the image plane. The DigiTizer chart was used to determine the accuracy of the camera calibration which was reported in mean colour difference. A large value of mean colour difference suggests a failure of the calibration. The image capturing only can be carried out when the calibration was succeeded. The images captured via this system were saved in both tiff format and nef (raw) format.

3.2.1.1 Camera characterisation

The digital images from the camera contain the RGB colour information which is device dependent. To display the same colour on a monitor, the camera RGB values have to be first transformed into a device independent colour

space, such as CIE XYZ. A polynomial regression transformation is a prediction model to estimate the related CIEXYZ values from camera RGB values. In the present study, this method was chosen to characterise the camera. The polynomial model was fully explained in Section 2.4.1.

As described in Section 2.4.1, the training samples and the polynomial model order can affect the model accuracy. Therefore, different training sample sets and polynomial models with different orders were investigated before deciding on the required characterisation model. For the training sample sets, three sets of colour data were used, including 24 colour samples from the *Macbeth Color Checker Chart* (MCCC); 110 skin colours from the *Pantone SkinTone™ Guide*; 92 skin colours from *Spectromatch Ltd*. silicon skin colour chart. The image RGB values of these colour samples and their XYZ tristimulus values were obtained via the Nikon D7000 camera and TSR, respectively. The silicon samples were measured using the SP, as the sample size was too small to measure with the TSR. In these training sample sets, the MCCC was frequently used to build and to evaluate a characterisation model (see Section 2.4.1). Two skin colour sample sets were selected for use as previous research has suggested that the colour of the training samples should be similar to that of the target samples to increase the model accuracy (Cheung, 2004). The impact of the number of training samples was investigated by using two skin colour sample sets together to train the polynomial model. Sixteen real skin colour samples included the RGB and the CIE XYZ from Nikon D7000 camera and the TSR, respectively, were used as a test sample set to evaluate the accuracy of the model. These 16 skin colour samples were from eight participants, including two facial locations from four ethnic groups of both genders. Figure 3.2.1.2 (a)-(b) show the distribution of the colours in 3 training sample sets and the test sample set in CIELAB $L^*C^*_{ab}$and CIELAB a^*b^* planes.



(a) (b)

Figure 3.2.1.2 The range of the training samples (silicon samples in blue; *Pantone SkinTone™ Guide* colour samples in grey; *Macbeth Colour Checker Chart* samples in yellow) and test samples (test sample in orange) in (a) in CIELAB $\boldsymbol{L^*C^*_{ab}}$ plane; (b) in CIE $\boldsymbol{a^*b^*}$ plane.

In these figures, the MCCC sample set covered the largest area in $L^*C^*_{ab}$ and a^*b^* planes. It however, has a limited number of colour patches that are close to skin colour. In $L^*C^*_{ab}$ plane, the silicon sample set covered larger area compared to that of the Pantone training sample set. Both training sample sets have good coverage compared to the test sample set in a^*b^* plane. In the polynomial model test, these two skin colour sample sets were used together, and separately, to train the polynomial models.

Five polynomial models with different terms were tested. These five models are listed in Table 3.2.1 and they were the models proposed and tested by Hong, *et al*. (2001) and Cheung, *et al*. (2004).

Table 3.2.1 The tested polynomial models with different number of terms

3 terms	R,G,B
4 terms	R,G,B,1
6 terms	R,G,B,RG,RB,GB
10 terms	R,G,B,RG,RB,GB,R^2,G^2,B^2,1
20 terms	R,G,B,RG,RB,GB,R^2,G^2,B^2,RGB,R^2G,G^2B,B^2R,R^2B,G^2R,B^2G,R^3,B^3,G^3,1

Each polynomial model was trained by four training sample sets. In total, twenty models (5 polynomial models×4 training sample sets=20 models) was evaluated via the data from 16 real skin colour samples (RGB from the digital camera, CIEXYZ based on the measurements of PR650). The procedure of evaluating the characterisation accuracy is shown in Figure 3.2.1.3.

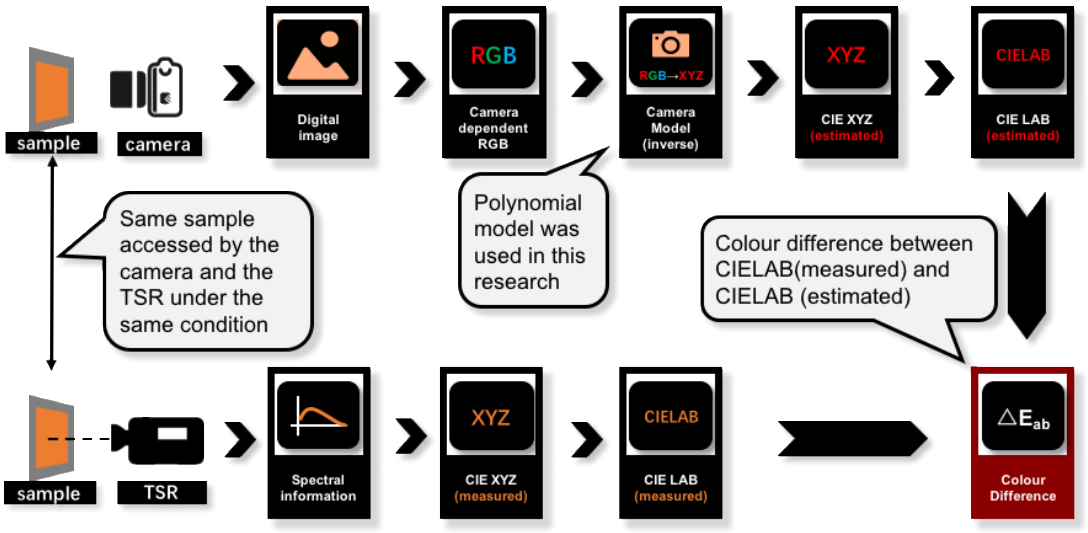


Figure 3.2.1.3 The procedure of the accuracy test of the camera's characterisation model.

As shown in this figure, the RGB and the CIE XYZ of test samples were measured via the same camera and TSR, respectively. Through the camera characterisation model, the estimated results can be obtained. The colour difference between the estimated results and TSR measured result was used to evaluate the accuracy of the model. Small colour difference indicated the

good accuracy of the model. Figure 3.2.1.4 shows the test results. The average colour difference of the two locations and the 8 participants was used to evaluate the model accuracy

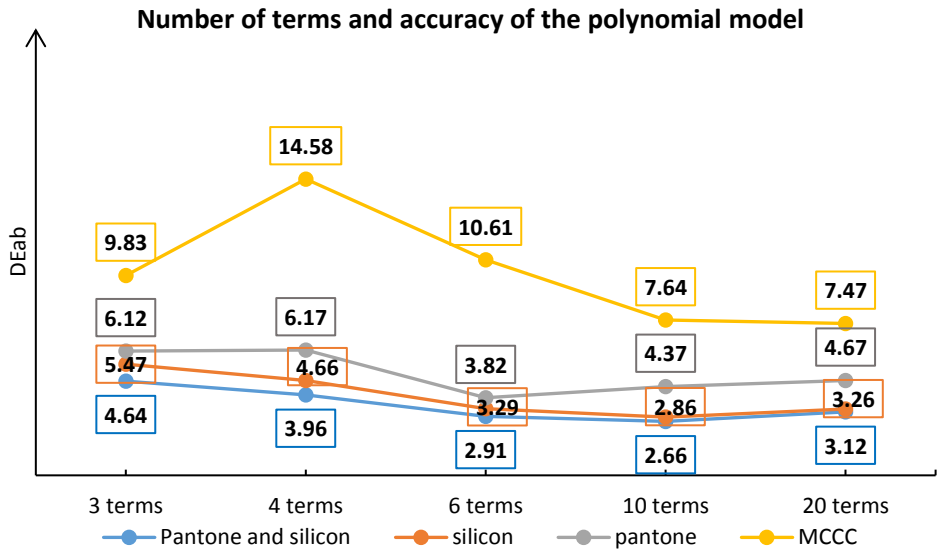


Figure 3.2.1.4 The test results of using four training sample sets to train five polynomial models.

From Figure 3.2.1.4, it can be seen that the training sample set with both Pantone and silicon samples gave the most accurate predictions among the four training sets. For this training sample set, the polynomial model with 10 terms gave the most accurate prediction amongst all five polynomial models, with a mean colour difference of 2.66 ΔE^*_{ab}. The silicon training sample set gave similar test results compared to the combined Pantone and silicon training sample set. The most accurate prediction of this training set was with 10 terms and gave a colour difference of 2.86 ΔE^*_{ab}. In these four training sets, the MCCC gave the least accurate prediction results.

In conclusion, the training sample set with the best coverage in terms of skin colour area gave a more accurate estimation. Therefore, it was decided that the combined Pantone and silicon training set would be used to train a polynomial model with 10 terms to characterise the digital camera.

3.2.1.2 The short-term repeatability of the facial image system

The short-term repeatability of the *DigiEye facial image system* was investigated via continuous capture colour chart and real human images. Here, the MCCC chart was selected and was captured five times continuously and the colours (RGB values) of the two skin colour patches, the light skin tone and dark skin tone patches, on the MCCC were extracted via Matlab. Also, seven participants (two Caucasians, four Chinese and one South Asian) had their two facial locations (forehead and cheekbone) marked and had their facial images captured continuously four times. The RGB values of the marked locations were extracted via Matlab. Then the camera characterisation model, which mentioned in the last section, was used to transform these RGB values into CIEXYZ coordinates. Then these CIE XYZ tristimulus values were transformed into CIELAB values. The MCDM of two colour patches of the five continuously captured MCCC images and of the two facial locations of the four continuously captured facial images were used to determine the short-term repeatability of the facial imaging system. Table 3.2.2 lists the mean MCDM value of the images of the MCCC chart and of the images of seven human participants.

Table 3.2.2 The mean MCDM values of the MCCC chart and the human participants.

	MCCC chart	Human subjects
Mean MCDM	0.05	0.54

It can be seen from the mean MCDM values in the table that less stable at capturing real human images than that of capturing a colour chart. This can be caused by the human participants cannot stay as still as the colour chart in the light booth. Overall, the imaging system had a satisfying short-term repeatability.

3.2.2 Monitor

An EIZO® LCD (ColorEdge CG243W) monitor was used in the psychophysical experiment of the present study. The psychophysical experiment was carried out in Zhejiang University, China. The PR650 which was used to characterise the camera was not available on this site. So, another TSR, *Jeti Specbos 1201* (Jeti), was used to evaluate the colourimetric performance and build a characterisation model for the monitor. Here, it is assumed that these two instruments gave similar measurement results. The Jeti was placed 60 cm away from the monitor to make measurements, as shown in Figure 3.2.2.1. This distance was the same as the observing distance in the psychophysical experiment. The colour gamma, brightness and the correlated colour temperature of the monitor were adjusted through the control panel of the monitor by following the manufacturer's instructions. The colour gamma was set to have a value of 2.2. The luminance of a displayed white was set to have a value of 100 cd/m^2. The correlated colour temperature was set to be 6500 K. To determine these characteristics of the monitor, except for the warm-up time test, all the evaluation tests were carried out after the monitor had been allowed to warm up for approximately 90 mins.

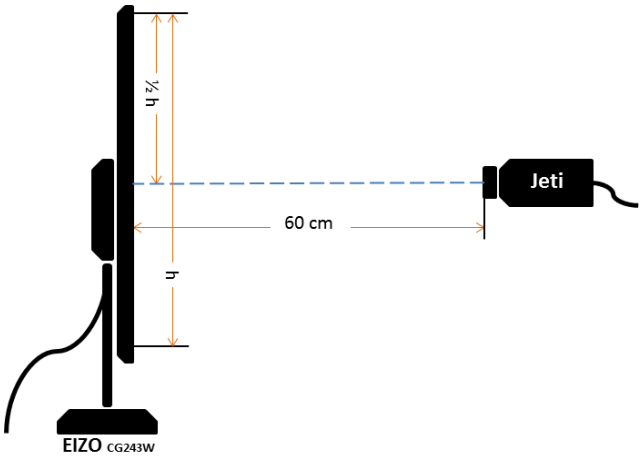


Figure 3.2.2.1 The monitor measurement set-up.

3.2.2.1 Monitor warm up

The warm-up time of the monitor was investigated by using the Jeti to measure luminance at the centre of the monitor every minute from the turn-on time for a period of 100 minutes. A white colour patch was displayed at the centre of the monitor. The variation in this measure is shown in Figure 3.2.2.2. It can be seen from this figure that the monitor starts to give a constant output of approximately 100 cd/m^2 after being turned on for approximately 65 mins. This indicated that the monitor needed at least 65 mins to warm up and reach a stable output.

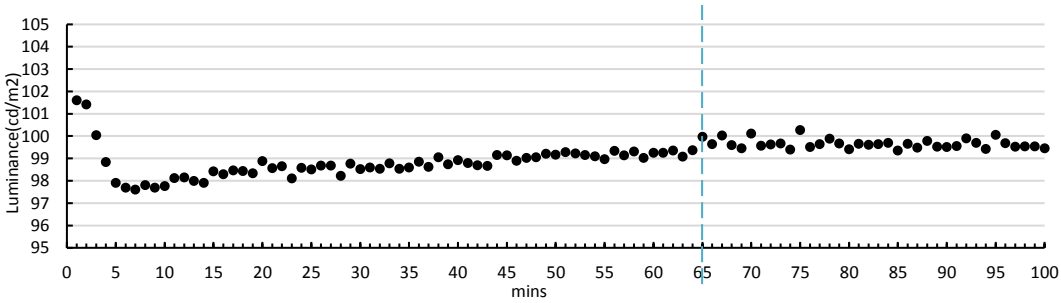


Figure 3.2.2.2 The variation of the luminance of the monitor from 0 minutes to 100 minutes.

Figure 3.2.2.3 shows the variation in the correlated colour temperature (CCT) during the warm-up process. From this figure, it can be seen that the CCT starts to become stable after the monitor had been turned on for approximately 85 mins. The CCT of the monitor after this warm-up period was approximately 6470 K. It can be seen from these two figures that the monitor needed at least 85 mins to warm up.

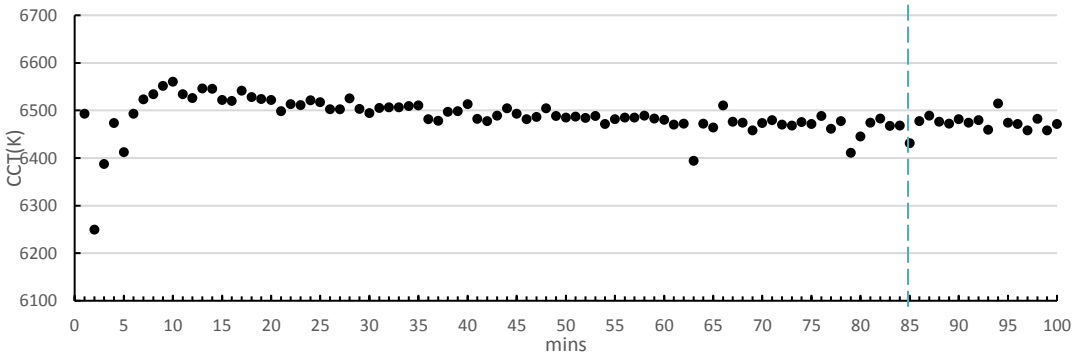


Figure 3.2.2.3 The variation of the CCT of the monitor from 0 minutes to 100 minutes

3.2.2.2 The luminance, the CCT and gamma of the monitor

From Section 3.2.2.1, it can be seen that the luminance and the CCT of the monitor stabilised after a warm-up period to have values of 100 cd/m^2 and 6470 k, respectively and that these values were close to the values set via the control panel of the monitor.

The colour gamma of the monitor can be calculated using the formula shown in Equation 3.2.

$$\gamma = \frac{log(V_{out})}{log(V_{in})} \qquad \textbf{Equation 3.2}$$

where V_{out} is the output luminance of the monitor for one of the channels. V_{in} is the digital input to the same channel. In this equation, the monitor black level is removed.

Eighteen levels for each channel was displayed and measured by using the setup that is shown in Figure 3.2.2.1. The measurement results are listed in Tables A1 to A5 in Appendix A. These results are plotted in Figure 3.2.2.4.

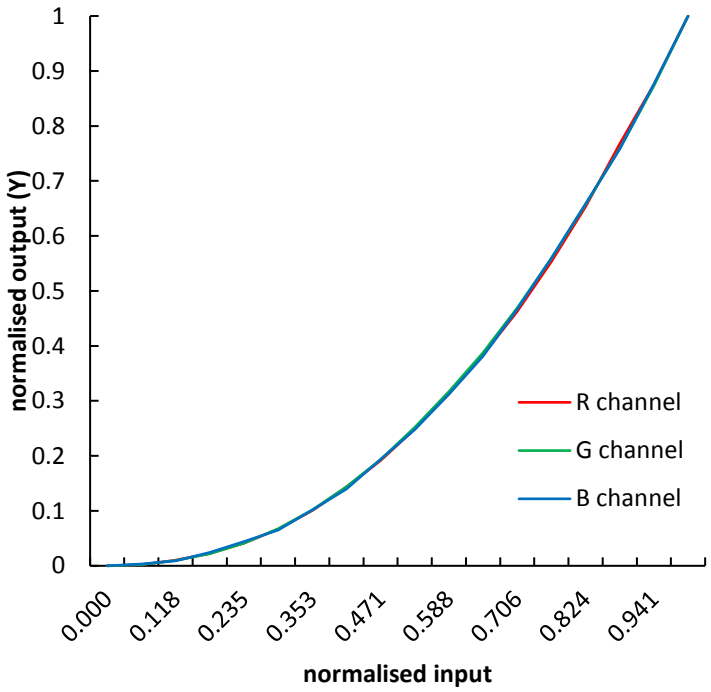


Figure 3.2.2.4 The relationship of the input value and the output value of the monitor used in the present study.

The gamma of the red, green and blue channels can be obtained using Equation 3.2, to give values of 2.19, 2.18 and 2.18, respectively. These gamma values were very close to the value set through the control panel of the monitor.

3.2.2.3 Monitor spatial uniformity

Nine (3×3) points on the screen were measured to determine the spatial uniformity of the monitor, as shown in Figure 3.2.2.5.

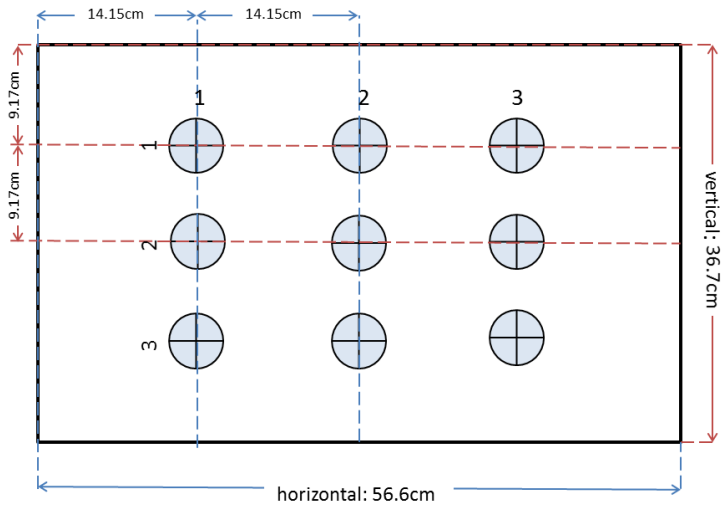


Figure 3.2.2.5 Measuring plan for the uniformity test.

The Jeti was used to measure these 9 spots and the measurement results were transformed into CIELAB values for further analysis. The ΔE^*_{ab} values between each measuring point and the mean of these nine points are plotted in the Figure 3.2.2.6 It can be seen in this figure that the variation alone the horizontal direction was more consistent than that in the vertical direction. The MCDM value was also calculated to determine the overall uniformity of the monitor and had a value of 1.06 ΔE^*_{ab}. The manufacturer's specification for the screen uniformity of this monitor is $\Delta E^*_{ab} \leq 3$: this indicated that this monitor was in a good condition for the experiment (Eizoglobal, 2012).

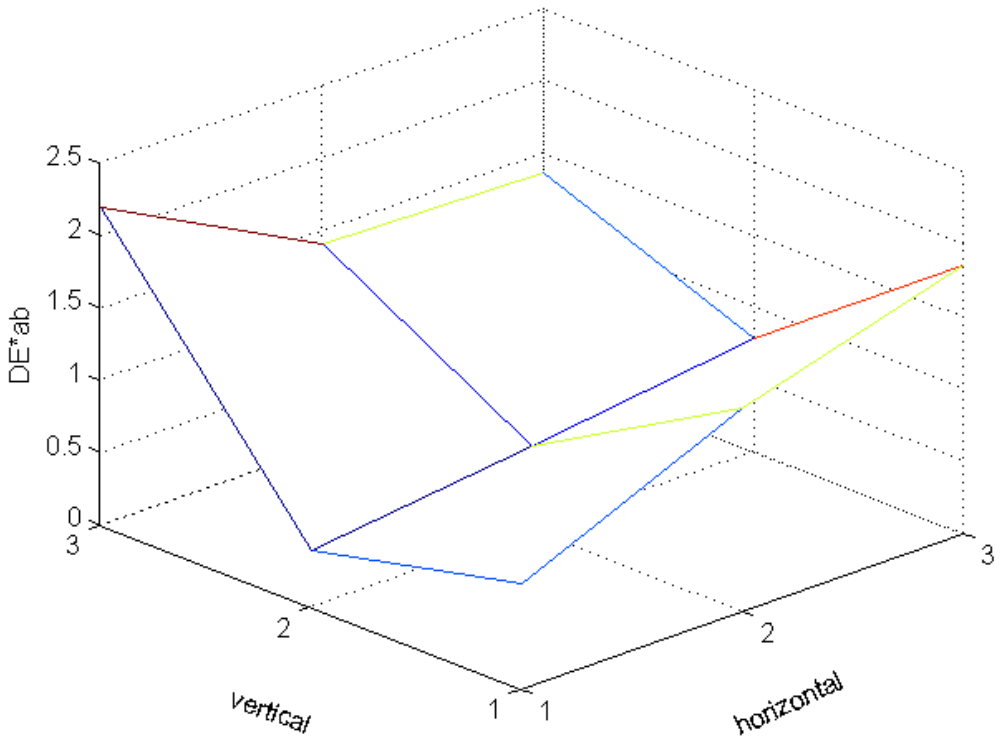


Figure 3.2.2.6 The measurement results of the uniformity test.

3.2.2.4 Repeatability

The repeatability of the monitor was monitored to determine its stability. Three tests, including short-term, medium-term and long-term repeatability tests, were conducted as part of these tests.

Short-term

The short-term repeatability investigated the performance of the monitor within a single time measurement. This was evaluated by measuring five colours (red, green, yellow, blue and white) at the centre of the monitor, each five times continuously. The MCDM of the five repeat measurements of each colour were calculated, and then the overall mean MCDM value was

calculated for evaluating the short-term repeatability. Table 3.2.3 lists the results of the short-term repeatability test.

Medium-term

To determine the medium-term repeatability, five measurements were made throughout a day at 3 hr intervals. For each measurement, five colours (red, green, yellow, blue and white), which were displayed at the centre of the monitor, were measured using the Jeti. The measurement results were transformed into CIELAB, then the MCDM of these five measurements was calculated. The MCDM value was used to determine the medium-term repeatability, as listed in Table 3.2.3.

Long-term

To determine the long-term repeatability, five measurements were made throughout the psychophysical experiment period (about two months). The same five colours were measured by the Jeti. The MCDM of each colour and the mean MCDM of the five colours were calculated to evaluate the long-term repeatability of the monitor, as listed in Table 3.2.3.

Table 3.2.3 The monitor repeatability test results.

	Short-term	Medium-term	Long-term
MCDM	0.23	0.47	0.49

It can be seen from these results that the average variation of the monitor was less than 0.5 ΔE^*_{ab}. The average variation of the short-term repeatability was about half the medium-term value and the long-term variation. The average variation of the medium-term and the long-term repeatability were similar. These results imply that the monitor had good stability.

3.2.2.5 Monitor characterisation

3.2.2.5.1 Colorimetry of the monitor

Channel independence and colorant constancy are two key constituents of the colorimetry of the monitor for monitor characterisation. Many monitor characterisation models, such as 3DLUT, PLCC and GOG, have been proposed, based on these two properties (Sharma and Bala, 2002; Hainich and Bimber, 2016). The channel independence is a measure of the additivity of the similarity of the colour channels of the monitor. This property indicates that the displayed colour is only related to the independent single colour channels, the R, G and B channels in this case. The location of the pixel and the other colour channels should have no impact on the output colour. This means that the output of a colour (R_1, G_1, B_1) is equal to the sum of the outputs of the three channels (R_1, 0, 0), (0, G_1, 0) and (0, 0, B_1). The colorant constancy is the property of the monitor that the chromaticity of a single channel remains the same when the intensity of the input varied.

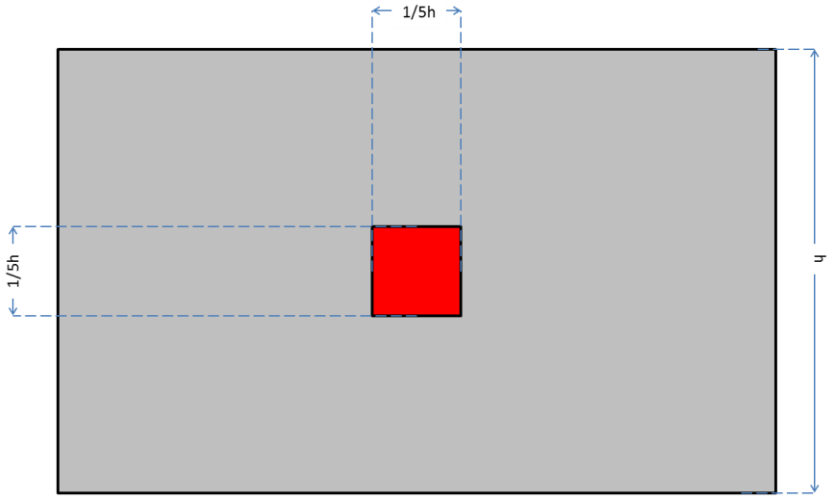


Figure 3.2.2.7 The interface of colour consistency test.

In this work, these two properties were evaluated by measuring the red, green, blue channels from 0 to 255 with the intervals of 15 digit levels, and 17 neutral colours from (0,0,0) to (255, 255, 255) with intervals of 15 digital levels. The measurement location was at the centre of the screen with a grey background (128,128,128), as shown in Figure 3.2.2.7. The measurement results were saved in CIE XYZ (2-degree) with the black offset removed. These measurement results were then transformed into CIELAB (D65/2) coordinates.

The channel independence was evaluated via the ΔE^*_{ab} values of the measurement results (CIELAB values of the neutral colours) and the calculated result (sum of the CIELAB values of the related R, G and B colours). The results for the 17 samples are shown in Figure 3.2.2.8

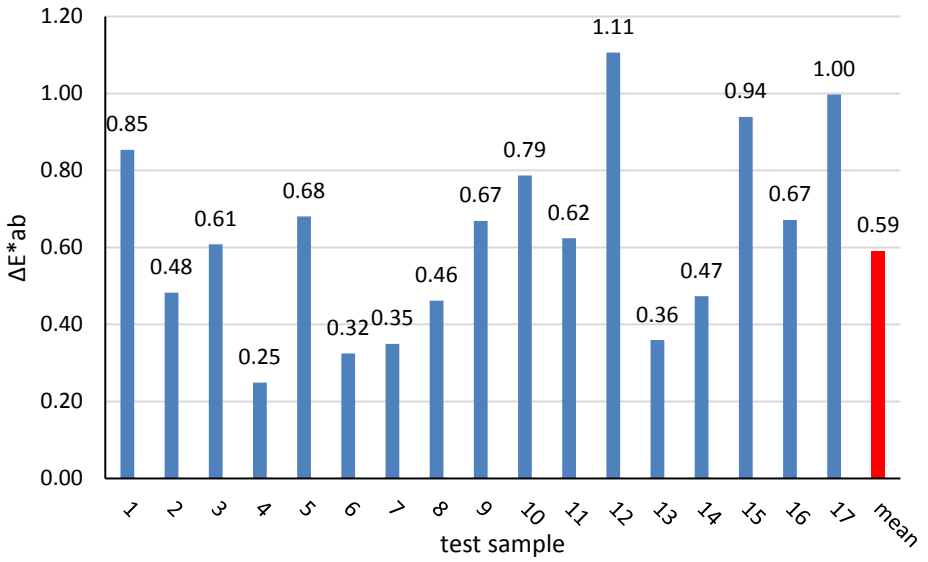


Figure 3.2.2.8 The results of the channel independence test.

It can be seen from the figure above that the mean colour difference between the measured results and calculated results were also calculated and the mean colour difference was 0.59 ΔE^*_{ab}. The channel independence is acceptable.

The colorant constancy was investigated by plotting the results in 1976 *u'v'* chromaticity diagram, as shown in Figure 3.2.2.9. In this figure, the

chromaticity shift of the three channels of the monitor appeared when the monitor input RGB values were (15,0,0), (0,15,0) and (0,0,15). Overall the RGB values remained consistent when the intensity increased.

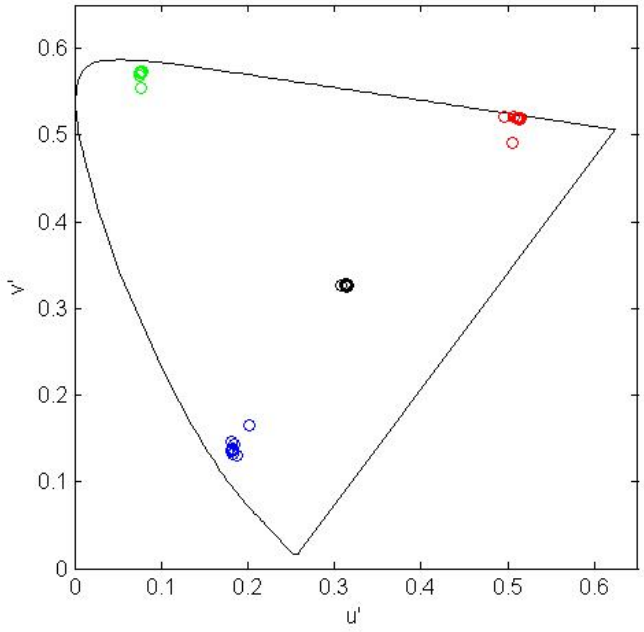


Figure 3.2.2.9 Colour consistency of the monitor

3.2.2.5.2 Monitor characterisation model

In the present study, the GOG model was chosen to characterise the monitor. The GOG model was fully explained in Section 2.4.2. Three GOG models were built for the R, G and B channels respectively. The accuracy of the models was evaluated by using skin colour samples and 24 colours of the MCCC.

Gained Offset Gamma model

Eighteen shades of red, green and blue, from 0 to 255 with an interval of 15 were used to train the GOG model for the R, G and B channel respectively. Figure 3.2.2.10 shows the procedure of to train the model.

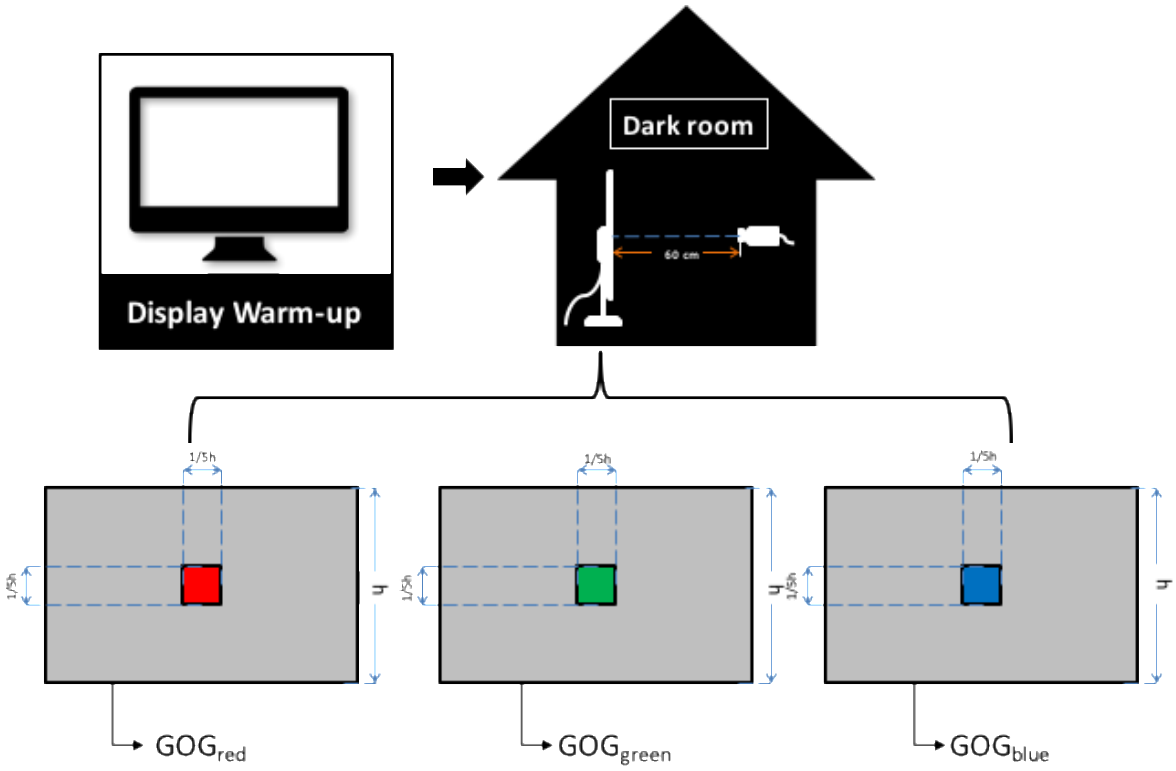


Figure 3.2.2.10 Procedure of building GOG model.

As illustrated in Figure 3.2.2.10, the monitor was first allowed to warm up, then the Jeti was used to measure the colour which was displayed at the centre of the screen, with a measurement distance of 60 cm. The measurement carried out in a dark room. The measurement results for the red, green and blue channels were used to train the GOG model for each channel. Through these three models, the CIE XYZ tristimulus values can be transformed into the monitor dependent R, G, B values, or the inverse.

Accuracy of the model

The procedure for evaluating the monitor characterisation model is shown in Figure 3.2.2.11.

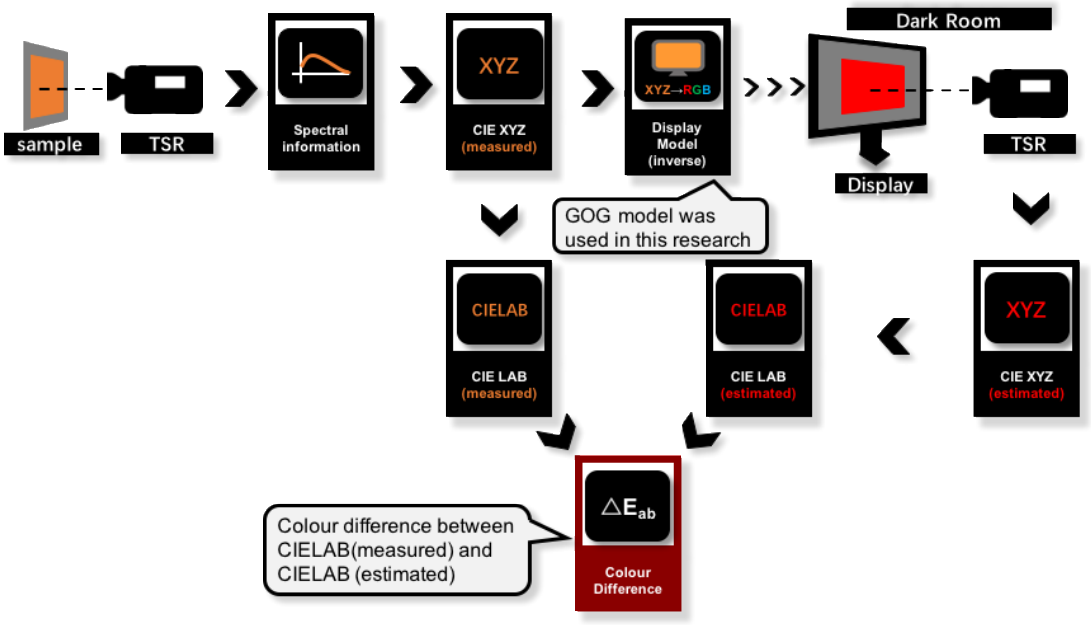


Figure 3.2.2.11 The procedure of the accuracy test of the characterisation model of the monitor.

As shown in the figure, these test samples were measured in the *DigiEye light booth* by using the PR650. These measurement results were then transformed into CIE XYZ values. Through the GOG model, the XYZ values were mapped into monitor dependent RGB values. These mapped colours were then displayed at the centre of the monitor with a neutral grey surround. The Jeti was used to measure these displayed colours, again with a measuring distance of 60 cm. The colour differences between these measured results of the reproduced colours on the monitor and of the original colour samples were calculated for evaluating the colour reproduction.

Twenty-four colours from the MCCC and seventeen skin colours from the *Pantone SkinTone™ Guide* were used to test the accuracy of the model.

These forty-one (24 +17) colour samples were measured in the DigiEye light booth by using the PR650. The results are listed in Table 3.2.4. The reason for the use of two different TSR was noted in the section above.

Table 3.2.4 The results of the accuracy test.

	all samples (ΔE^*_{ab})	*Pantone SkinTone™ Guide* (ΔE^*_{ab})	MCCC (ΔE^*_{ab})
Mean	1.18	1.20	1.16
Max	2.11	1.46	2.11
Medium	1.14	1.24	1.09

The mean colour difference of all forty-one test colours was 1.18 ΔE^*_{ab}. For the skin colour samples, the mean colour difference was 1.20 ΔE^*_{ab}. For the 24 MCCC samples, the mean colour difference was 1.16 ΔE^*_{ab}. The test results showed that the accuracy of the colour reproduction was acceptable.

The PLCC model was used to characterise an LCD monitor of the same brand as the one used here (Georgoula, 2015). The 24 MCCC samples were used to test the model accuracy. The mean colour difference between the original CIE XYZ values and the modelled CIE XYZ values was 1.8 ΔE^*_{ab} which compares very favourably with the value of 1.16 ΔE^*_{ab} calculated here, Table 3.2.4.

3.3 Summary of Chapter 3

In this chapter, the repeatability and colour shift of skin colour measurements was investigated for two instruments, a spectrophotometer (SP) and a tele-spectroradiometer (TSR), with four sets of different settings, three sets for the SP and one set for the TSR. The current analysis provides useful guidance for defining a protocol for skin colour measurements that allows for the comparison of different data sets obtained with different measurement instruments. Skin colour measurements for 11 subjects each with two facial areas, forehead and cheekbone, were performed and the variability was evaluated for different parameters. For the short-term repeatability, it can be seen from the MCDM values that different measuring field sizes and pressures applied while measurement was made do not affect the short-term repeatability of both instruments. However, different repeat measurement methods can affect the short-term repeatability greatly. Within two repeat measurement methods, the CT method has higher short-term repeatability than the CS method. The CT method is recommended to be used when using spectrophotometers. Different measuring distances and measuring field sizes were found to have the most significant colour shifts and colour differences for both instruments.

Through these repeatability tests, the measurement protocols of the PR650 and the CM700d were decided, and are considered in the following chapter. For the PR650, as the short-term repeatability between the two measurement positions was similar: position 1 was selected. For the CM700d, the MAVLP mask was selected, as it has a similar measurement field size as the PR650 with a measurement distance of 57.5 cm. The CM700d is a portable instrument, so ten locations were measured using this instrument. Apart from the CM700d and the PR650, the protocols for visual assessment and camera image capture are also decided and considered in detail in the next chapter.

The evaluation tests and the characterisation accuracy test of the monitor, which was used in the psychophysical experiment, were described in this chapter. Through the evaluation test, the warm-up time of the monitor was determined to be approximately 1.5 hr; the CCT of the monitor after warm-up was approximately 6470 K; the luminance of the monitor after warm-up was approximately 100 cd/m^2; the repeatability and the uniformity of the monitor were found to be acceptable. The channel independency and colour constancy were also evaluated and determined to be acceptable. The MCCC and skin colour patches were used to determine the accuracy of the monitor calibration model. The mean colour difference between the CIELAB values of the test colours and those of the colours that were displayed on the monitor was 1.16 ΔE^*_{ab} a commendably low result.

Chapter 4 Skin colour database collection

Skin colour measurements have been widely accumulated for skin colour reproduction on the different media, including paper, monitor and mobile phone screen (ISO, 2003). Many skin colour databases were built in the past, such as the SOCS database and Kawasaki database (Tajima *et al*., 1998; ISO, 2003), but these databases often included spectral data with a restricted range. This Chapter describes a database that was built to include measurements of skin colour from people of different ethnicity and from different locations on their bodies. The data were also measured using different methods. This experiment is referred as Experiment 1 in the thesis and is entirely devoted to the accumulation of skin colour measurements. The new skin colour database, to be designated the Leeds-Liverpool skin colour (LLSC) database, aims to form a major part of the CIE TC1-92 *Skin Colour Database* work. Three different measurement methods, including the use of a tele-spectroradiometer (TSR), a spectrophotometer (SP) and a set of skin colour charts (VA), were used to measure the skin colour of subjects at different locations. In addition, a facial imaging system (included a digital single-lens reflex (DSLR) camera) was used to capture facial images using defined measurement and capture protocols. In this chapter, the skin colour data in this database was analysed in terms of different measurement tools, subject ethnicities, subject genders and measurement locations on the body.

4.1 Experiment 1 Physical Measurement of Skin Colour

4.1.1 Experimental design

4.1.1.1 Measurement protocol

Three measurement tools, including a TSR, a SP and a set of skin colour charts (VA), were used to measure skin colour. In addition, a facial imaging system (included a digital camera) was used to capture facial images. Here named the facial imaging system as camera. The experimental protocols for using the two instruments, the TSR and the SP, which were the PR650 and CM700d, were determined by the tests described in Section 3.1.2 and they are listed in Table 4.1.1.

Table 4.1.1 The measuring protocols.

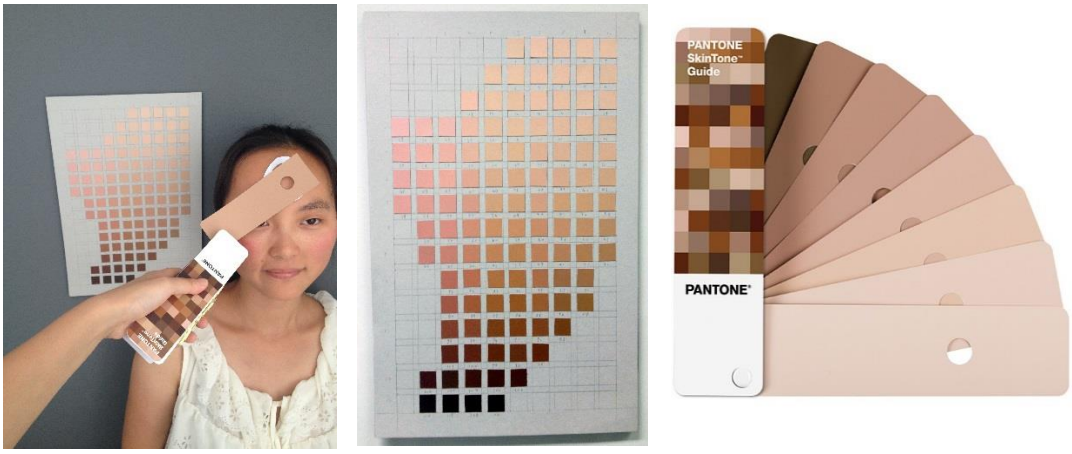
Instrument	Protocol
PR650	At position 1 (57.5cm from instrument to the subject); measuring 5 locations, forehead (FH), cheekbone (CB), cheek (CH), neck, back of hand (BH)
CM700d	Using the MAVL mask to measure 10 locations, forehead (FH), cheekbone (CB), cheek (CH), neck, the back of the hand (BH), inner forearm (IFA), outer forearm (OFA), chin, nose tip (NT), fingertip (FT)
Camera	4 images, a front face and a side face image with markers, a front face and a side face image without markers.
Visual assessment	3 trained observers, measuring 4 locations (forehead (FH), cheekbone (CB), cheek (CH), neck).

As noted in the table, facial images with markers and without markers were both captured. The images with markers were used to record the measurement locations. The RGB values at these measurement locations were extracted from the images without markers (the images with markers

were used as a reference). The visual assessment procedure was designed to enable the observer to find the best matching colour patch of the skin colour chart rapidly. The procedure is described in Section 4.1.1.2 and the protocol is also listed in Table 4.1.1.

4.1.1.2 Visual assessment

The visual assessment (VA) method was designed to ask observers to find a patch from a skin colour chart which was most closely matched the skin colour at a target location. In the present study, the *PANTONE SkinTone™ Guide* was used to measure skin colour inside a *DigiEye facial image* light booth. The light booth provided D65 illumination for achieving the best colour matching results using the *SkinTone™ Guide*. The colour chart includes 110 skin colours patches. Each of the colour patches was given a unique serial number, which coded the degree of yellowness/redness and lightness. But, with this serial number, it was difficult for observers to find the best matching colour efficiently from the *SkinTone™ Guide*. So, the colours were rearranged according to their lightness and yellowness/redness planes. A part of each chart in the *SkinTone™ Guide* was cut and attached to a neutral background, as shown in in Figure 4.1.1.1 (b). With this colour chart, the observer can first determine the range of the colours that might match the target location. Then find the fan deck sample in the *SkinTone™ Guide* with the same serial number, and compare the colour from the fan deck and the target location via the hole on the fan deck. After judging the difference between the fan deck and the target location, in terms of lightness and yellowness/redness, the observer then selected another fan deck sample which might be closer to the skin colour of the target location. This process continued until the best matching fan deck sample was found.



(a) (b) (c)

Figure 4.1.1.1 The visual assessment: (a) visual assessment environment; (b) *PANTONE SkinTone™ Shade*; (c) *PANTONE SkinTone™ Guide*.

The colour of each fan deck was measured by PR650 inside the light booth with a measurement distance of 57.5 cm. The measurement results were transformed into CIELAB values and used as the colour of the fan deck sample. Three trained observers took part in the visual assessments with each participant individually seeing the subjects in a random order. They were trained to assess the skin colour of the subjects at the designated illuminating and viewing geometry and to be familiar with the locations and relations between the different colours in the fan deck. The visual assessment results were recorded and the average of the results of the three observers was used as the final result for the target location.

4.1.1.3 The experimental procedure

The experimental procedure is shown in Figure 4.1.1.2.

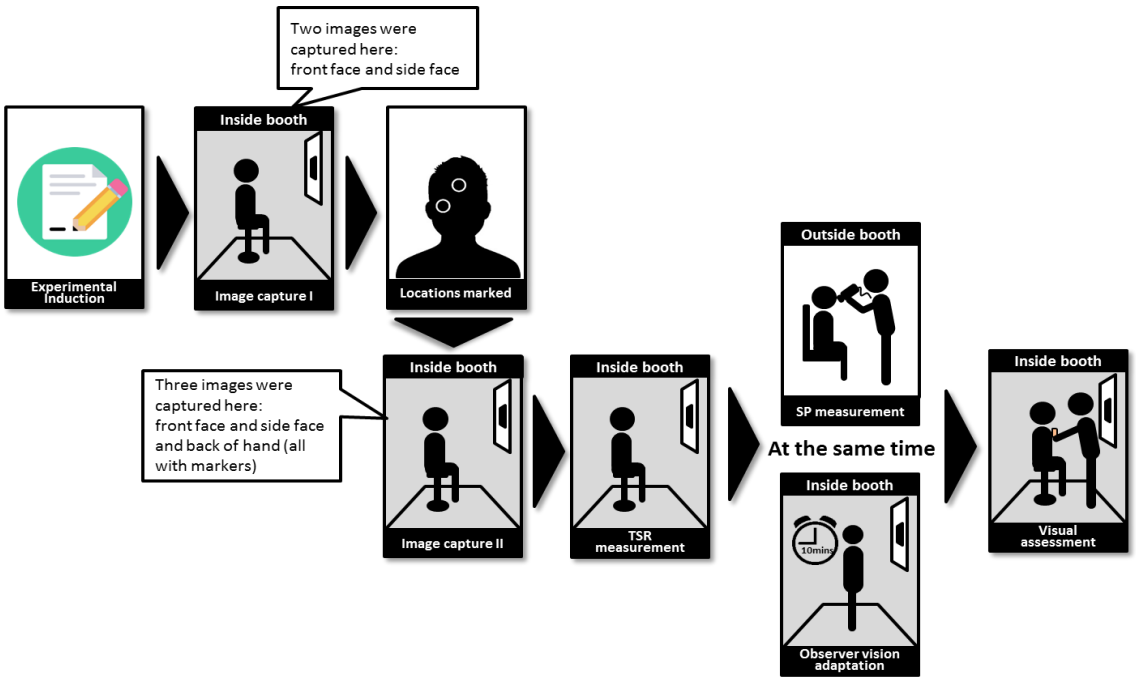


Figure 4.1.1.2 The experimental procedure of Experiment 1

As shown in Figure 4.1.1.2, the facial images without markers were captured first. The capturing distance was confirmed at every image capture. Then, the measurement locations were marked by the white circle tags which had identical size. After marking the measurement locations, the images with markers were captured at the same distance. After the image capture, the PR650 was used to measure the colour of the four facial locations. Each location was measured three times continuously. The measurement distance was checked before each measurement. The marked locations were then measured using the CM700d. Each location was also measured three times continuously. These repeat measurements of the PR650 and the CM700d were used to determine the short-term repeatability of these instruments for

the different measuring locations, different genders and different ethnicities. The average of these three repeat measurements was used as the final measurement result for the specified location. The visual assessment was carried out by three trained observers. Before the visual assessment, these observers stayed inside the light booth to adapt to the light environment for about 10 mins. The average CIELAB values of the colour chart that were selected by the three observers were used as the visual assessment result for this location.

4.1.2 Results and discussion

4.1.2.1 An overview of the LLSC database

All the accumulated data were included in the Leeds-Liverpool skin colour (LLSC) database which was established by Universities of Leeds and University of Liverpool. This database also forms an input to the work of CIE Technical Committee TC 1-92 *Skin Colour Database*. The LLSC database includes data from four ethnic groups, Caucasian, Oriental, South Asian and African, with both genders, as listed in Table 4.1.2.

Table 4.1.2 The information of the measured subjects

Ethnic group	Male	Female
Caucasian	14	65
Oriental	45	41
South Asian	7	6
African	5	5
Total	71	117

One hundred and eighty-eight people participated. Four different tools, including a PR650, a CM700d, a DSLR camera and a set of skin colour chart

(VA), were used to measure the skin colour of 10 locations and to capture facial images. These 10 locations included 6 facial locations, forehead (FH), cheekbone (CB), cheek (CH), nose tip (NT), chin and neck, and 4 body locations, which were the back of the hand (BH), inner forearm (IFA), outer forearm (OFA) and fingertip (FT). Most of the locations were marked by the same size markers. The four facial images included a front face and a side face image with and without markers. The RGB information of each location was extracted from the images without markers by reference the image with markers.

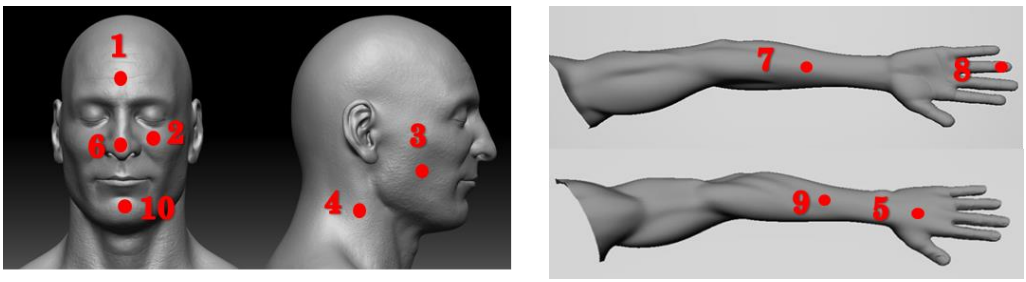


Figure 4.1.2.1 Ten measuring locations (1. forehead (FH), 2. cheekbone (CB), 3. cheek (CH), 4. neck, 5. the back of the hand (BH), 6. nose tip (NT), 7. inner forearm (IFA), 8 fingertip (FT), 9. outer forearm (OFA) and 10. Chin).

4.1.2.2 Repeatability of the instruments

The stability of PR650, CM700d and the imaging system were monitored by measuring two skin colour patches (the *light skin* and *dark skin*) of the Macbeth Colour Checker Chart (MCCC) under the same measurement conditions as the experiment. For the imaging system, the image of the MCCC chart was captured and the RGB values of the two skin colour patches on the MCCC chart were extracted via Matlab software. Then the camera characterisation model (see Section 3.2.1.1) was used to transform the RGB values into CIEXYZ values. Then the CIEXYZ values were transformed into CIELAB values. The MCDM of the repeat measurement results was used to determine the short-term, medium-term and long-term repeatability. The

MCDM of the PR650, CM700d and camera are listed in Table 4.1.3. The procedure of the short-term, medium-term and long-term repeatability tests are listed in Sections 4.1.2.2.1 to 4.1.2.2.3.

Table 4.1.3 The MCDM value for short-term, medium-term and long-term repeatability tests using colour patches.

	PR650	CM700d	Camera
Short-term	0.08	0.07	0.05
Medium-term	0.43	0.07	1.65
Long-term	0.65	0.13	1.75

4.1.2.2.1 The short-term repeatability

The short-term repeatability of the PR650 and CM700d were determined by measuring two skin colour patches of the MCCC 5 times continuously within 20 seconds. The MCDM value of these 5 repeat measurements was used to represent the short-term repeatability of these two instruments, as listed in Table 4.1.3. The short-term repeatability of the imaging system was explained in Section 3.2.1.2.

4.1.2.2.2 The medium-term repeatability

The medium-term repeatability of the PR650 and CM700d was evaluated through measuring the same two skin colour patches at five different times of the day with about 2 hrs intervals (with a period of 10hr). The MCDM value of these five measurements was used to represent the medium-term repeatability of the instruments, as listed in Table 4.1.3. Similarly, the camera captured the MCCC image at the same position at the same time. The CIELAB

values of two skin colour patches were extracted and the MCDM value of them was used to represent the medium-term repeatability of the camera.

4.1.2.2.3 The long-term repeatability

The measurements of these two skin colour patches on every experimental day, which was within a period of 70 days, were used to monitor the long-term repeatability of these two instruments. The MCDM of these measurements was used to represent the long-term repeatability, as listed in Table 4.1.3. The camera also captured the MCCC image on every experimental day. The colours of the two skin colour patches were extracted to monitor the long-term repeatability of the camera.

The data in Table 4.1.3 suggest that all three instruments had satisfactory short-term, medium-term and long-term repeatability. As might be expected, the short-term repeatability was relatively better than the others. The imaging system had larger MCDM values at medium term and long term than the CM700d and PR650. Comparing with the CM700d and the PR650, the consistency of the position of two colour patches in the digital image was less repeatable. This may cause the variation in the medium-term and long-term results.

4.1.2.2.4 Agreement between observers for visual assessment

The agreement between the visual assessment results of the 3 observers was evaluated by using the MCDM value. Table 4.1.4 lists the MCDM value for the two facial locations (forehead (FH) and cheek (CH)) of each participant. In this test, seven subjects, including 3 Caucasians (C), 3 Orientals and 1 South Asian (SA), were measured.

Table 4.1.4 The agreement between three observers (MCDM).

Subjects	FH (ΔE^*_{ab})	CH (ΔE^*_{ab})
Subject 1 (O)	3.86	2.64
Subject 2 (C)	1.33	2.13
Subject 3 (O)	3.94	3.82
Subject 4 (O)	5.66	4.43
Subject 5 (C)	0.57	2.31
Subject 6 (C)	3.23	2.67
Subject 7 (SA)	3.10	3.00
Mean	3.05	

The overall MCDM value of the 3 observers was 3.05 ΔE^*_{ab}. This value is similar to the visual assessment agreement results of Alfvin and Fairchild (1997) which is about 3.03 ΔE^*_{ab} (2˚). And it is larger than the variation of the measurement results of the instruments. The skin colour chart that was used to carry out the visual assessment contained 110 skin colours but the observers sometimes found that the skin colour of the target area could not be found in the charts. Under this circumstance, the observer was taught to select the sample that was closest to the target. This could certainly lead to an increase in the disagreement between the observers.

4.1.2.3 Comparing spectral reflectance between different ethnic groups and different instruments

Spectral reflectance can be considered as the fingerprint of a colour. i.e. the basic element of colour specification. It is for this reason that the LLSC database contains spectral data. In this section, the differences between the

results from the two measurement instruments are first reported for each of four ethnic groups. Figure 4.1.2.2 shows the two measurement results in terms of spectral reflectance for the two instruments. Each spectral reflectance function was an average of the measurements from four facial locations. The measurement results of PR650 and CM700d were plotted in blue and orange, respectively.

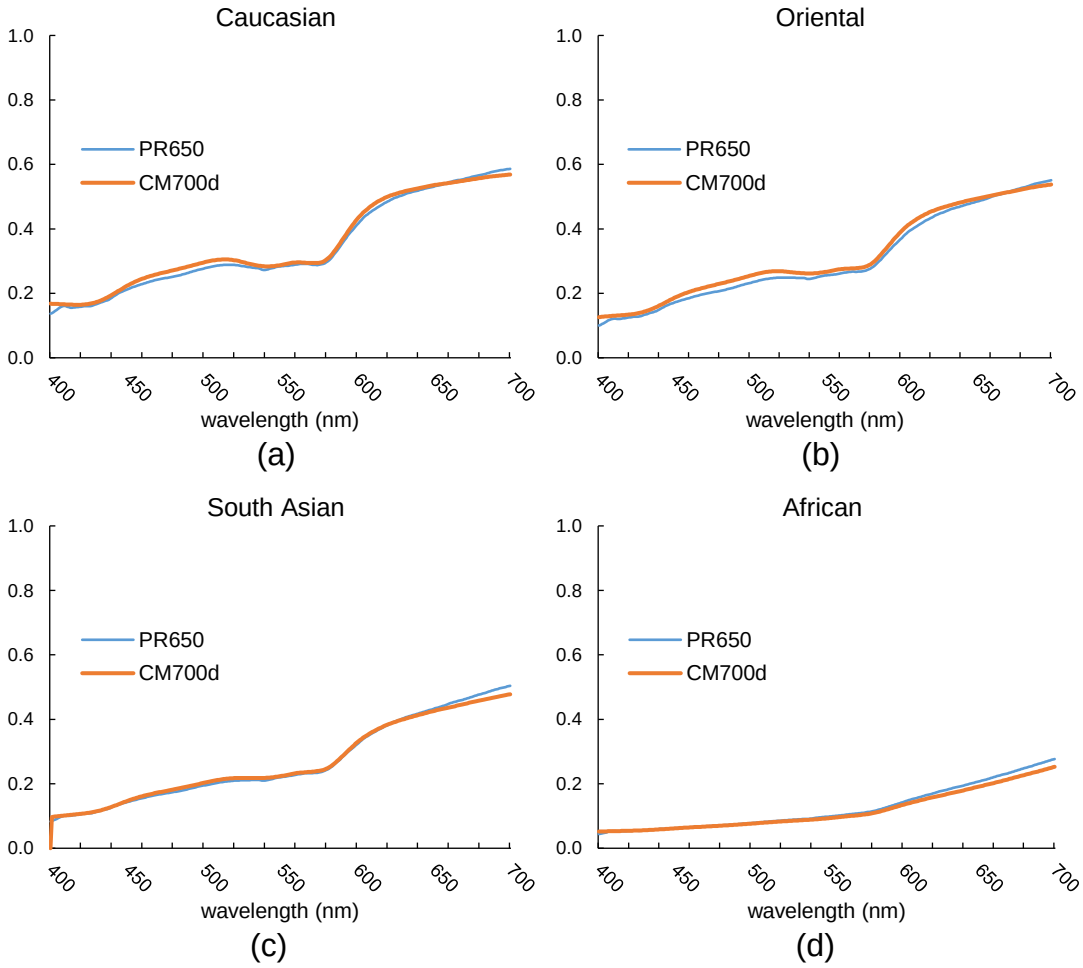


Figure 4.1.2.2 The comparison of spectral reflectance as measured by two instruments. (a) The measurement results of Caucasian subjects; (b) Oriental subjects; (c) South Asian subjects; (d) African subjects.

It can be seen from these figures that the measurement results of the two instruments had good agreement over the range between 400 nm and 600 nm. A systematic discrepancy can be found at above 660 nm, where the spectral reflectance as measured by the PR650 is always above that of the CM700d

instrument. This could be due to the different measurement method between the two instruments, i.e. they are non-contact and contact, respectively. The pressure cannot be avoided for the contact type of instrument. The pressure can cause dispersal of the blood at the measurement area. So, the result of the CM700d appeared reddish and have more absorption in the high wavelength range. The CM700d is a contact type of instrument. The pressure or stretch at the contact area might cause the measurement results appear reddish.

In Figure 3.1.1.3, the SPD of the light source contains two peaks at about 435 nm and 545 nm, as the fluorescent light tube is used as the light source. But no peak was found in the spectral reflectance, which were calculated by using the SPD shown in Figure 3.1.1.3, in Figure 4.1.2.2. This is because the peaks in the SPD were cancelled out after the calculation (see Appendix B for more detail).

The two reflectance in each subfigure of Figure 4.1.2.2 were transformed into CIELAB values (D65/2), then the colour difference between the two CIELAB values was calculated and is listed in Table 4.1.5.

Table 4.1.5 The colour difference between the two spectral reflectances in Figure 4.1.2.2.

Sub-figure	PR650vsCM700d (ΔE^*_{ab})
Caucasian (Figure 4.1.2.2 (a))	1.23
Oriental (Figure 4.1.2.2 (b))	1.78
South Asian (Figure 4.1.2.2 (c))	1.06
African (Figure 4.1.2.2 (d))	1.95

The colour differences between the two spectral reflectances in each subfigure were below 2 ΔE^*_{ab}. The colour difference between the spectral reflectance of the South Asian that measured via PR650 and that measured via CM700d was the smallest (1.06 ΔE^*_{ab}). The colour difference of the spectral reflectance of the African was the largest (1.95 ΔE^*_{ab}).

The differences between different ethnic groups were also investigated via the spectral reflectance, as shown in Figure 4.1.2.3. The Caucasian, Oriental, South Asian and African were marked in blue, orange, grey and yellow, respectively. The spectral reflectance of each ethnic group is the average reflectance measurement from four facial locations (FH, CB, CH and neck).

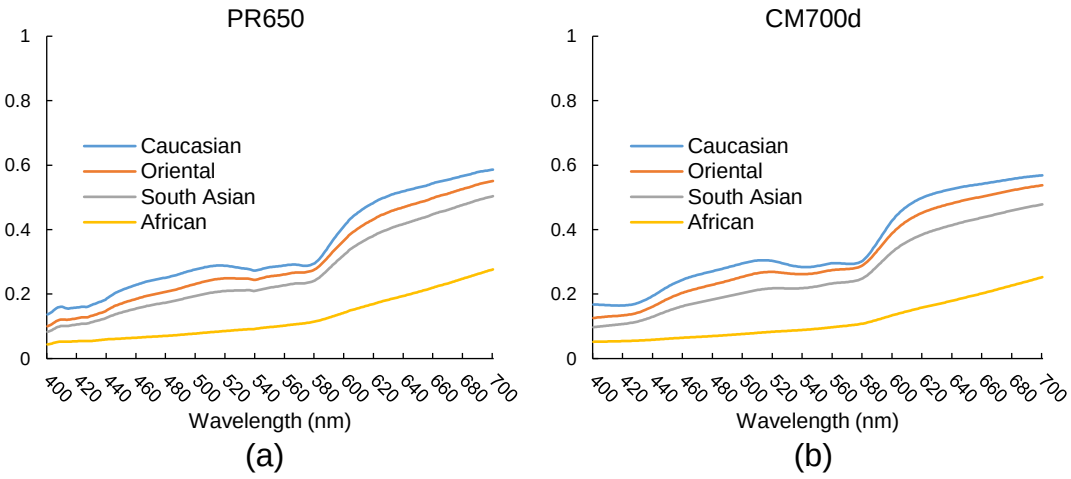


Figure 4.1.2.3 Comparison of the difference between the different ethnic groups using average spectral reflectance. (a) the measurement results of PR650; (b) the measurement results of CM700d.

Both instruments results showed a consistent trend that the Caucasian had the highest lightness, followed by Oriental, South Asian and African in that order. This is the same as the finding which gained from the plots in CIELAB $L^*C^*_{ab}$ plane, see Figure 4.1.2.9 (a).

The spectral reflectance of four ethnic groups in the two subfigures of Figure 4.1.2.3 were transformed into CIELAB values (D65/2). The colour difference between the CIELAB values of two ethnic groups was calculated and is listed in Table 4.1.6.

Table 4.1.6 The colour difference between the two ethnic groups.

	PR650 (ΔE^*_{ab}) (Figure 4.1.2.3 (a))	CM700d (ΔE^*_{ab}) (Figure 4.1.2.3 (b))
Oriental vs Caucasian	4.15	3.76
Oriental vs South Asian	3.61	4.45
Oriental vs African	21.40	23.84
Caucasian vs South Asian	7.17	7.63
Caucasian vs African	23.94	25.81
African vs South Asian	17.97	19.76

In Table 4.1.6, the colour differences between the Oriental and Caucasian groups and between Oriental and South Asian groups was about 4 ΔE^*_{ab}. The colour difference between the South Asian and African groups was about 18 $\Delta E^*_{ab.}$

4.1.2.4 The impact of the ethnicity and measurement locations on the short-term repeatability

The mean MCDM values of three repeat measurement of each location of each ethnic group are listed in Table 4.1.7 and Table 4.1.8, for CM700d and PR650, respectively. For each ethnic group, the averaged MCDM value of all locations was used to represent the repeatability of this ethnic group, as listed in the last column of each table. For each location, the MCDM value for all ethnic groups was averaged and used to represent the repeatability of this location, as in the last row of each table. The average MCDM of the MCDM of all the locations and ethnic groups is marked in yellow.

Table 4.1.7 The average MCDM values of different ethnic groups at different locations using CM700d (FH, forehead; CH, cheek; CB, cheekbone; BH, back of hand; IFA, inner forearm; OFA, outer forearm; NT, nose tip; FT, fingertip).

	FH	CB	CH	Neck	BH	5 locations mean	IFA	OFA	Chin	NT	FT	mean of Ethnic Group
Oriental	0.42	0.38	0.36	0.40	0.24	0.36	0.22	0.20	0.60	0.79	0.67	0.43
Caucasian	0.42	0.46	0.42	0.55	0.28	0.43	0.21	0.18	0.59	0.86	0.86	0.48
South Asian	0.40	0.76	0.34	0.41	0.26	0.43	0.13	0.19	0.46	0.63	0.77	0.44
African	0.17	0.18	0.23	0.23	0.19	0.20	0.09	0.11	0.24	0.29	0.75	0.25
Mean of locations	0.35	0.44	0.34	0.40	0.24	0.36	0.16	0.17	0.47	0.64	0.76	**0.40**

For the CM700d, the mean MCDM values of different locations (last row of the table) showed that the measurement results of the locations on the arm were lower than those on the face, except for the fingertip, probably because the locations on arm were relatively flat. For all the locations, the NT was least repeatable one (MCDM=0.76 ΔE^*_{ab}). For each ethnic group, the OFA and IFA were more repeatable and the FT was least repeatable than the others. The mean MCDM values of different ethnic groups (last column) showed that the African group were more consistent than the other three. The mean MCDM value for the African group was about 0.25 ΔE^*_{ab}, which was half of the mean MCDM of the other ethnic groups.

The average of the MCDM values of the five locations, which were measured via the PR650 and the CM700d, were listed in the column named 5 locations mean in Table 4.1.7. This averaged MCDM value was compared with the averaged MCDM values of the same five locations that measured via the PR650 as shown in Table 4.1.8.

Table 4.1.8 The average MCDM values of different ethnic groups at different locations using the PR650 (FH, forehead; CB, cheekbone; CH, cheek, BH, back of the hand).

	FH	CB	CH	Neck	BH	Mean of ethnic group
Oriental	0.33	0.76	0.87	0.98	0.63	0.71
Caucasian	0.47	0.80	1.04	1.13	1.01	0.89
South Asian	0.40	0.66	0.99	0.72	1.04	0.76
African	0.75	0.55	2.30	0.78	0.62	1.00
Mean of locations	0.49	0.69	1.30	0.90	0.82	**0.84**

For the PR650, the mean MCDM of different locations (in the last row of Table 4.1.8) showed that the FH was the most precise location while CH exhibits the largest variations. From the mean MCDM of each ethnic group (in the last column of Table 4.1.6), it can be concluded that the measurement results of the Oriental group and the African group were the most and least repeatable, respectively. For each ethnic group, the measurement results of the FH were more repeatable than the others, except for the African group which had the most repeatable measurement result at the location CB. The measurement results for the neck were less repeatable than the others for the Oriental and Caucasian group. For the South Asian group, the MCDM value for the locations CH and BH were about 1.00 ΔE^*_{ab}, which was twice of the average MCDM value for the FH location. For the African group, the measurement results for the CH location of the African group were the least repeatable amongst all the locations and all ethnic groups.

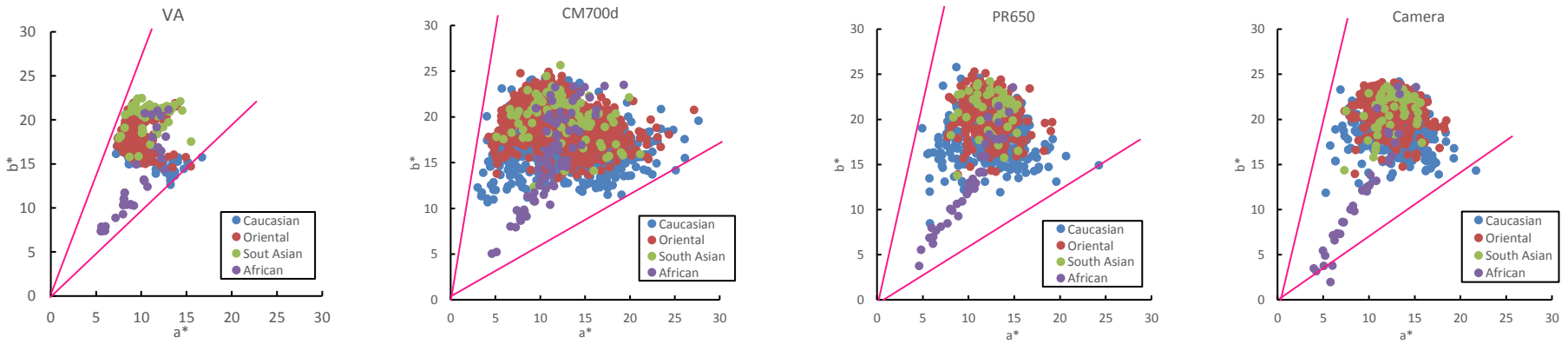
It can be seen from Table 4.1.7 and Table 4.1.8 that the averaged MCDM value of the CM700d measurements of the five locations (0.36 ΔE^*_{ab}) was about half of that of the PR650 (0.84 ΔE^*_{ab}). This indicates that the measurement results for the CM700d are more precise than those of the PR650. It is typical that non-contact measurement results are more variable than contact measurement results. In addition, the non-contact instrument, the

PR650, gave the worst repeatable results (1.00 ΔE^*_{ab}) for the African group and the contact instrument, the CM700d, gave the most repeatable results (0.25 ΔE^*_{ab}). This can be caused by the measuring area being fixed for the CM700d, as it contacted the measurement location. But, the measuring area cannot be precisely located for the PR650, because it is a non-contact instrument. Therefore, the non-contact instrument has poorer short-term repeatability than the contact instrument. For the African group, the measurement time can be longer than the other groups, because the integration time of the instrument is longer when measuring dark colours. Any shift of the measurement area can caused by movement of the participant and can result in the measurement being less repeatable.

4.1.2.5 Colour distribution of different tools

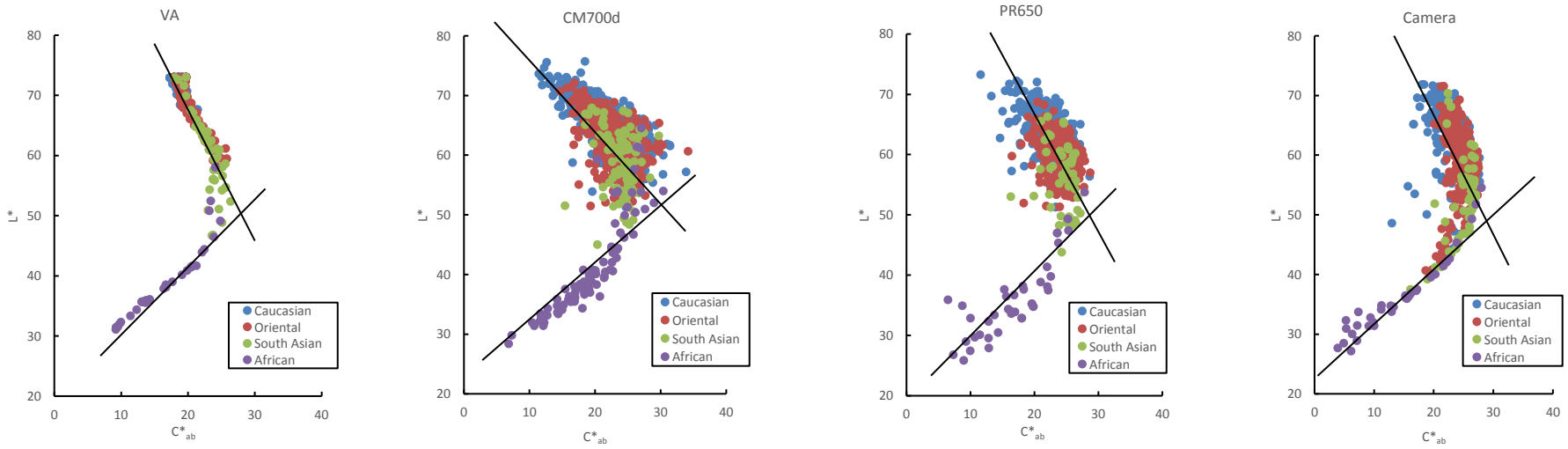
All the measurement results were transformed into CIEXYZ tristimulus values and then to CIELAB (D65/2) coordinates for further analysis. For the camera, the average RGB values of the marked locations were mapped into CIEXYZ space using the polynomial model. Then the CIEXYZ values were transformed into CIELAB values. Here, the camera was not used as a measurement instrument. The transformed results were only plotted in this section to get a concept of the skin colour distribution.

The measurement results of all 188 participants of the from visual assessment (VA), the PR650, the CM700d and the camera were plotted in CIELAB a^*b^* and $L^*C^*_{ab}$ planes, as shown in Figure 4.1.2.4 and Figure 4.1.2.5, respectively. In these figures, Figure 4.1.2.4 (a) and Figure 4.1.2.5 (a) illustrate the measurement results for four locations (FH, CH, CB and neck) from four ethnic groups measured by using VA; Figure 4.1.2.4 (b) and Figure 4.1.2.5 (b) illustrate the measurement results of ten locations (FH, CH, CB, neck, chin, nose, BH, IFA, OFA and FT) from four ethnic groups measured by the CM700d; Figure 4.1.2.4 (c) and Figure 4.1.2.5 (c) show the measurement results of five locations (FH, CH, CB, neck and BH) from four ethnic groups measured using the PR650; Figure 4.1.2.4 (d) and Figure 4.1.2.5 (d) show the CIELAB coordinates of four locations (FH, CH, CB, neck) from four ethnic groups that was gained from the camera images. In these figures, different ethnic groups were coded in different colours. The Caucasian, Oriental, South Asian and African were marked in blue, red, green and purple, respectively. The black lines in $L^*C^*_{ab}$ plane were drawn to show the trend of the data distribution.



(a) (b) (c) (d)

Figure 4.1.2.4 The distribution of the measurement results of (a) visual assessment; (b) CM700d; (c) PR650; (d) Camera; in $\boldsymbol{a^*b^*}$ plane.



(a) (b) (c) (d)

Figure 4.1.2.5 The distribution of the measurement results of visual assessment (a), CM700d(b), PR650(c) and camera (d) in $L^*C^*_{ab}$.

It can be seen from Figure 4.1.2.4 and Figure 4.1.2.5 that the measurement results that were obtained via the different methods had similar distributions in a^*b^* and $L^*C^*_{ab}$ planes. In a^*b^* plane, the Caucasian, Oriental and South Asians skin colours of different subjects have a larger variation in hue angle than the African, as shown in Figure 4.1.2.4. For the African, the measurement results mainly varied in chroma.

The mean, maximum, minimum and range of the hue angle of each ethnic group in Figure 4.1.2.4 are listed in Table 4.1.9. The range of the hue angle was calculated as the maximum value minus the minimum value. The hue angle of the different ethnic groups is coded with different colours: the yellow, red, blue and green represent the Oriental, Caucasian, South Asian and African group, respectively.

Table 4.1.9 The mean, maximum, minimum and the range of the hue angle of each ethnic group and of each measurement method.

Method	Ethnicity	Mean	Max.	Min.	Range
VA	Oriental	62.80 ˚	68.15 ˚	43.57 ˚	**24.58 ˚**
	Caucasian	59.68 ˚	67.33 ˚	43.31 ˚	**24.02 ˚**
	South Asian	61.97 ˚	68.15 ˚	48.39 ˚	**19.75 ˚**
	African	53.71 ˚	63.44 ˚	48.32 ˚	**15.12 ˚**
CM700d	Oriental	58.57 ˚	76.52 ˚	34.83 ˚	**41.69 ˚**
	Caucasian	54.53 ˚	78.54 ˚	29.40 ˚	**49.14 ˚**
	South Asian	58.09 ˚	73.85 ˚	36.98 ˚	**36.87 ˚**
	African	53.28 ˚	65.49 ˚	43.00 ˚	**22.48 ˚**
PR650	Oriental	58.18 ˚	73.48 ˚	33.41 ˚	**40.07 ˚**
	Caucasian	56.05 ˚	75.41 ˚	31.96 ˚	**43.44 ˚**
	South Asian	59.20 ˚	68.71 ˚	46.96 ˚	**21.75 ˚**
	African	51.83 ˚	59.02 ˚	43.07 ˚	**15.95 ˚**

Patterns can be found from this table that the average hue angle of the Oriental is higher than that of the Caucasian Group for the measurement results of all three methods. Also, the average hue angle of the African group is the smallest. This indicates that the Oriental group had yellower skin colour than the Caucasian group. The African group had the reddest skin of all. For

the results from the PR650 and the CM700d, the hue angle range of the Caucasian group is slightly larger than that of the Oriental group. This finding is agreed with the finding of Xiao *et al*. (2016) (see section 2.2.2). For the VA, the hue angle ranges of the Oriental and Caucasian groups were similar. And they were about half of the hue angle ranges that of the PR650 and the CM700d. This can due to the hue angle range of the skin colour chart that used in VA is limited (see Figure 3.2.1.2). For all three methods, the African group had the smallest hue angle range. This shows that the hue angle variation of the African group's skin colour was smaller than that of the other ethnic groups.

It can be found from the above that all four methods shown similar distribution. Here, the measurement results from PR650 is used as an example to show the variation between subjects. The average of four locations, including FH, CH, CB and Neck, on the face of each subject was calculated to represent the skin colour of his/hers face. The plots of the facial skin colour of all 188 subjects were showed in Appendix B, Figure B9. The maximum and minimum value of the L^*, C^*_{ab} and h_{ab} of each ethnicity were listed in Table 4.1.10. The range of these three scales were calculated by using the maximum value minus the minimum value, as listed in Table 4.1.10.

Table 4.1.10 The Maximum, minimum and the range of the L^*, C^*_{ab} and h_{ab} of four ethnicities.

		L^*	C^*_{ab}	h_{ab} (degree)
Caucasian	**Max.**	70.30	25.54	63.68
	Min.	55.80	15.61	43.30
	range	**14.50**	**9.93**	**20.38**
Oriental	**Max.**	65.52	26.76	63.89
	Min.	53.24	20.14	52.36
	range	**12.28**	**6.62**	**11.54**
South Asian	**Max.**	64.58	26.31	64.33
	Min.	49.14	21.67	53.59
	range	**15.44**	**4.64**	**10.74**
African	**Max.**	48.99	24.36	57.45
	Min.	28.87	9.75	49.03
	range	**20.12**	**14.61**	**8.42**

It can find from the data shown in Table 4.1.10 that the range of the $L^*$and the C^*_{ab} values of the African subjects is the largest. The range of the h_{ab} of the African subjects is the smallest. These implies that the variation of the African subjects is mainly in L^* and C^*_{ab}. And they have limited variation in h_{ab}. Note here, only the data of 10 African subjects were included. The range of C^*_{ab} of the South Asian is the smallest of all four ethnicities. The range of the h_{ab} of the South Asian subjects are similar to the Oriental subjects. Note here, the data of 13 South Asian subjects were included. The range of the h_{ab} of the Caucasian is the largest. The ranges of the L^* and C^*_{ab} are similar to the Oriental subjects. These showed that the Caucasian subjects have relatively larger difference in hue angle.

In $L^*C^*_{ab}$plane, the distributions of the measurement results from four methods were similar to the curve of a banana, as shown in Figure 4.1.2.5 (a)-(d). The Caucasian and Oriental data distributed alone a negative correlation line. The African group distributed alone a positive correlation line. The South Asian group results, which were measured using the PR650 and the CM700d, showed limited variation in chroma and mainly varied in lightness. But their

measurement results from the VA and the Camera images were distributed alone a positive correlation line and negative correlation line, respectively. These can be summarised as the L^* and the C^*_{ab} of the skin colour of the Caucasian and Oriental groups had the negative association which means that the skin colour of higher lightness value is lower in chroma value. The L^* and the C^*_{ab} of the skin colour of the African group had positive association, which means that the skin colour of the African group with higher lightness value is also higher in chroma value.

4.1.2.6 Skin whiteness and blackness trend line

In Section 4.1.2.5, Figure 4.1.2.4 showed the distribution of the skin colour of the Caucasian, Oriental and African groups were alone the negative and positive correlation lines in CIELAB $L^*C^*_{ab}$ plane. In the previous study, the skin colours of Japanese female with higher lightness value and lower chroma value was found to be appeared whiter (Yoshikawa *et al*., 2001). Adopting this finding with the finding that was gained in Section 4.1.2.5, the variation of the skin whiteness or skin blackness might be able to be described via the negative and positive trend line. To further investigate this, the linearly fitted trend lines of the PR650 data for each ethnic group were generated and are plotted in Figure 4.1.2.6. The trend line of each ethnic group in Figure 4.1.2.6 was generated by using the trend line function in *Microsoft Office Excel 2013*.

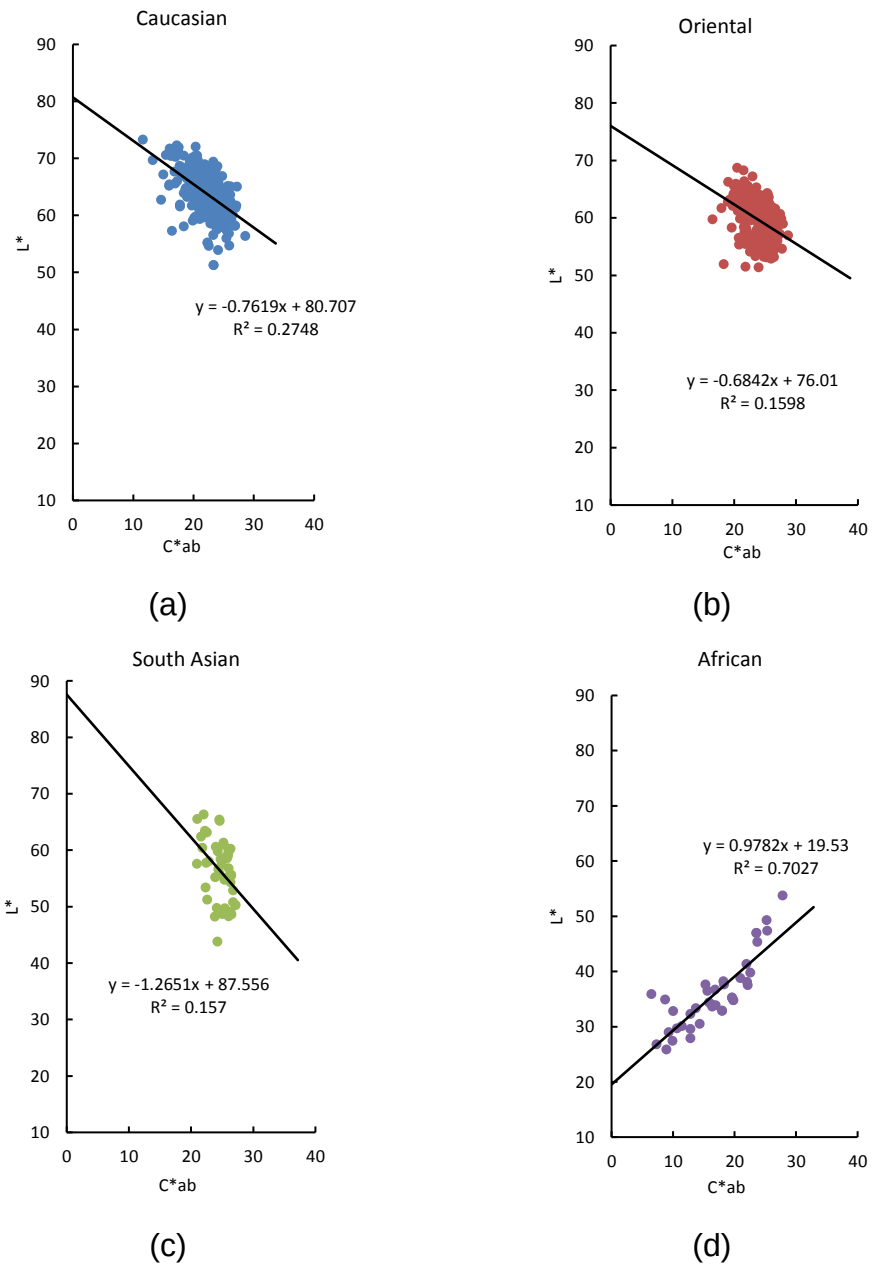


Figure 4.1.2.6 The distribution of the data for the four ethnic groups in $\mathbf{L^*C^*_{ab}}$ plane (measured by the TSR) with the associated trend line (y is the L^*; x is the C^*_{ab}). (a) Caucasian group; (b) Oriental group; (c) South Asian group; (d) African group.

In this figure, the skin colour data were distributed alone the generated trend line, except the South Asian group whose skin colours mainly varied in lightness. The South Asian group is not discussed here. This implies that the

trend lines could be used to illustrate the gradual variation of the skin colour in terms of skin whiteness or blackness. The intercept on the L* axis of the trend lines indicates can be understood as the estimated whitest/ blackest skin colour of the ethnic group. Here, these intercepts on the L* axis designated the *Skin Whiteness coefficient (W)* and *Skin Blackness coefficient* (bk) for the data of negative association and the positive association, respectively. Adopted from the formulae of Berns (2014), the skin whiteness/blackness can be represented via the distance between the skin colour and the *Skin Whiteness/Blackness coefficient* in $L^*C^*_{ab}$ plane, and designated the *Skin Whiteness Index* and the *Skin Blackness Index*. Equation 4.1 shows the formula used to calculate the *Skin Whiteness Index* (W_s).

$$W_s = 100 - \sqrt{(W - L^*_i)^2 + (C^*_{ab_i})^2} \qquad \textbf{Equation 4.1}$$

where L^*_i and $C^*_{ab_i}$ are the lightness and chroma value of subject i. W is the *Skin Whiteness coefficient.* The values of W for the Caucasian and the Oriental groups are listed in Table 4.1.11.

The *Skin Blackness Index* (BK_s) can be calculated via the blackness coefficient, as shown in Equation 4.2.

$$BK_s = \sqrt{(L^*_j - bk)^2 + (C^*_{ab_j})^2} \qquad \textbf{Equation 4.2}$$

where L^*_j and $C^*_{ab_j}$ are the lightness and chroma values of a subject j, respectively. bk is the *Skin Blackness Coefficient*. The values of bk for the African group was listed in Table 4.1.11.

Note here, the *Skin Whiteness Index* (W_s) can only be used to indicate the skin whiteness of the Caucasian and Oriental group. The *Skin Blackness*

Index (BK_s) can only be used to indicate the skin blackness of the African group.

Table 4.1.11 The *Skin Whiteness coefficients* (for Caucasian and Oriental) and *Skin Blackness coefficient* (for African) from different measuring tools.

Tools / Ethnic group	PR	CM	VA
Caucasian (W)	80.7	88.5	100.0 (102.8)*
Oriental (W)	76.0	88.8	100.0 (111.3)*
African (bk)	19.5	16.7	19.5

when the intercept of the L is greater than 100 (as indicated in the bracket), a value of 100 was used as the maximum whiteness coefficient.

From Equation 4.1 and Equation 4.2, it can be seen that a larger the value of W_s^* indicates a whiter skin colour and smaller value of small BK^* indicates a blacker skin colour. The skin whiteness and blackness index proposed in the present study was compared with the Individual Typology Angle (ITA˚) which is a scale that proposed to determine the 'lightness' of the skin colour in the previous study (see Section 2.3 for details). The PR650 data of the LLSC database were used for this comparison. These skin colour data were transformed into skin whiteness/blackness values and the ITA°. These skin whiteness/blackness values are plotted against the value of ITA° for different ethnic groups in Figure 4.1.2.7. The linearly fitted lines (by using *Microsoft Excel*) are also shown in these figures. The r^2 values listed in the figure indicate the degree of fit of the line to the data.

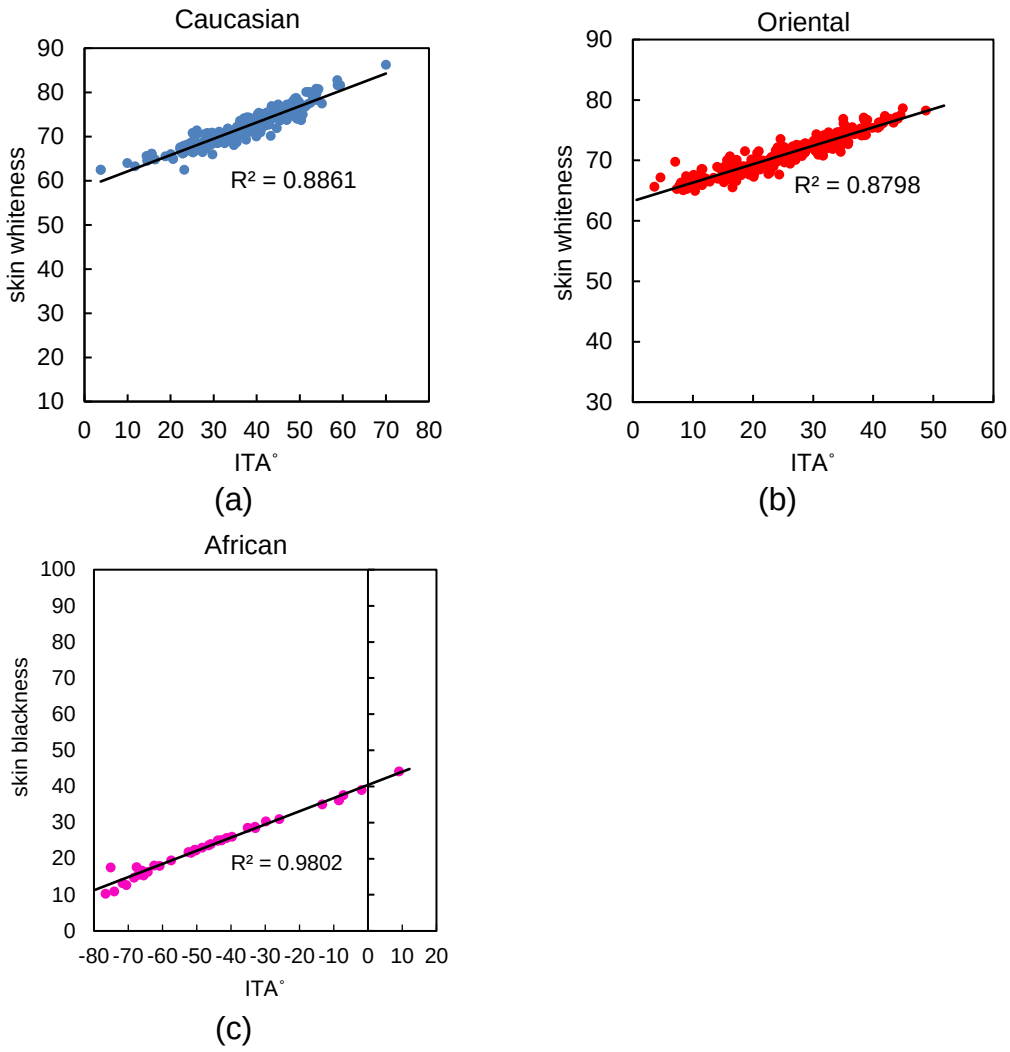


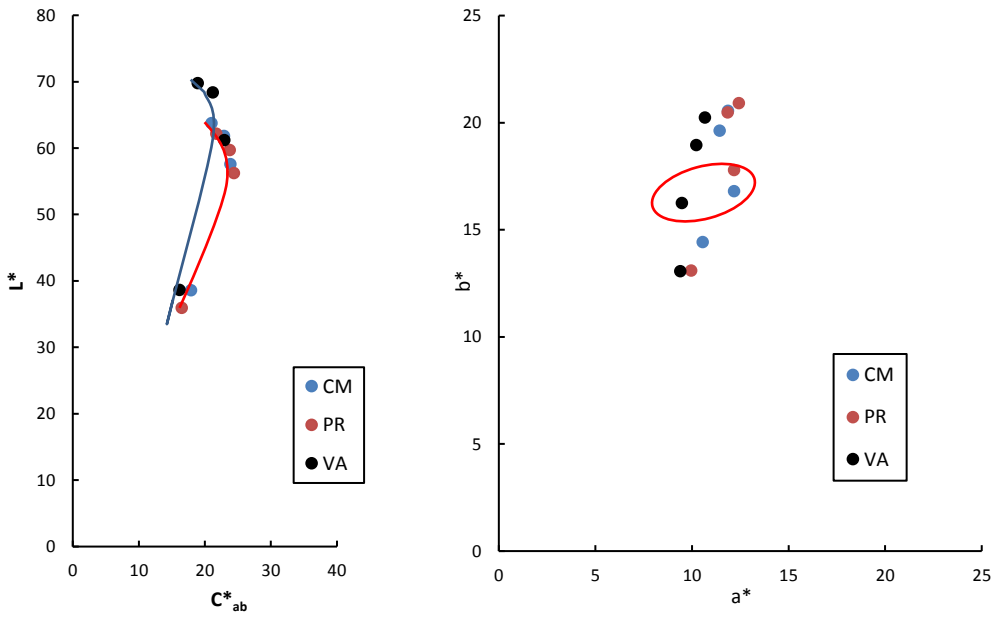
Figure 4.1.2.7 The agreement between the skin whiteness/blackness index and the ITA°. (a) Caucasian group's data; (b) Oriental group's data; (c) African group's data.

These figures show that the skin whiteness and the ITA° have a positive linear correlation. This indicates that the skin whiteness index and skin blackness index are valid scales to represent the skin whiteness and blackness. As mentioned in Section 2.3, the ITA° was calculated from the CIELAB L^* and b^* values, which means a^* information is not included. The skin whiteness and blackness indices were calculated from the CIELAB L^* and C^*_{ab} values, which means these indices included the L^*, a^* and b^* information. In the past research, a^* was found closely correlated to the erythema index and melanin

concentration (Takiwaki, Ovengaard and Serup, 1994; Alaluf *et al*., 2002; Shrive and Parra, 2000; Takiwaki *et al*. 2002). And a^* also be found related to the different locations of the human subjects (Xiao *et al*., 2017). This indicates that the skin whiteness and blackness indices include more colour information than the ITA°.

4.1.2.7 Comparing different ethnic groups and measurement methods

The differences of the colour distributions between ethnic groups and between different measuring methods were investigated via the mean CIELAB values of the participants of each ethnic group. For each method, the CIELAB values of the four locations (FH, CH, CB and neck) of each ethnic group were averaged and represented as a dot, as shown in Figure 4.1.2.8. So, four dots were included in each method, representing the Caucasian, Oriental, South Asian and African group. To investigate the difference between different tools, the dots of the same method were coded with the same colour, as shown in Figure 4.1.2.8. The blue curve and the red curve in Figure 4.1.2.8 (a) linked the data from VA and CM700d/PR650, respectively. To investigate the difference between the different ethnic groups, the dots of the same ethnic group were coded with the same colour, as shown in Figure 4.1.2.9, where the red circle in Figure 4.1.2.8 and Figure 4.1.2.9 marked out the large difference between measurement methods.



(a) (b)

Figure 4.1.2.8 The average CIELAB values measured by the CM700d (CM), PR650 (PR) and skin colour charts (VA) in (a) $\mathbf{L^*C^*_{ab}}$ plane and (b) $\mathbf{a^*b^*}$ plane.

It can be seen from Figure 4.1.2.8 that the results from the CM700d and the PR650 are similar. The results from the CM700d are slightly higher in lightness and slightly lower in chroma. The hue angle of results from the two instruments agreed well. The measurement results of VA are higher in lightness and appear yellower than the results from the other two instruments.

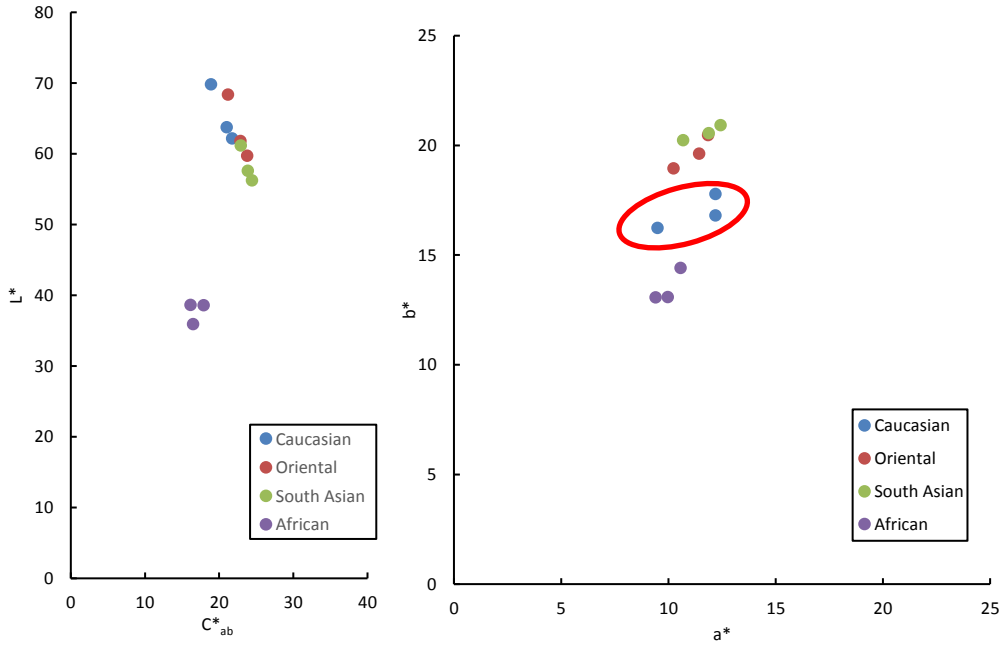


Figure 4.1.2.9 The average CIELAB values of four ethnic groups (Caucasian: blue; Oriental: red; South Asian: green; African purple) measured by using the CM700d, PR650 and skin colour charts in: (a) $\mathbf{L^*C^*_{ab}}$ plane and (b) $\mathbf{a^*b^*}$plane.

In Figure 4.1.2.9, consistent trends can be found by comparing the results between ethnic groups. In $L^*C^*_{ab}$ plane, the Caucasian group had the highest lightness followed by the Oriental, South Asian and African in that order, as shown in Figure 4.1.2.9 (a). In Figure 4.1.2.9 (b), the South Asian group had the highest chroma followed by the Oriental, Caucasian and African group in that order. Therefore, the finding can also be concluded as the Caucasian skin colour appeared the whitest, followed by the Oriental skin colour, the South Asians with the Africans skin colour as the darkest.

The red circle in Figure 4.1.2.8 (b) and in Figure 4.1.2.9 (b) marks the disagreement in the results. It can be seen from these figures that these were the measurement results of the Caucasian group of different measurement

methods. As shown in these figures, the measurement results from VA were yellower than the results from the CM700d and PR650.

4.1.2.8 Comparing different measurement locations

4.1.2.8.1 Difference between facial locations

For each ethnic group, the skin colour of each measurement location was represented by the average of the CIELAB values of all the participants within this ethnic group. These mean CIELAB values of the four facial locations (forehead (FH), cheekbone (CB), cheek (CH) and neck) of four ethnic groups were plotted in CIELAB $L^*C^*_{ab}$ plane and CIELAB a^*b^*plane to evaluate the colour difference between facial locations. The measurement results of two instruments, the PR650 (PR) and the CM700d (CM) were plotted in $L^*C^*_{ab}$ plane and a^*b^* plane, as shown in Figure 4.1.2.10 and Figure 4.1.2.11, respectively. The measurement results of VA were plotted in $L^*C^*_{ab}$ plane and a^*b^* plane, as shown in Figure 4.1.2.12. The blue lines were drawn in a^*b^*plane to mark the hue angle range of the data.

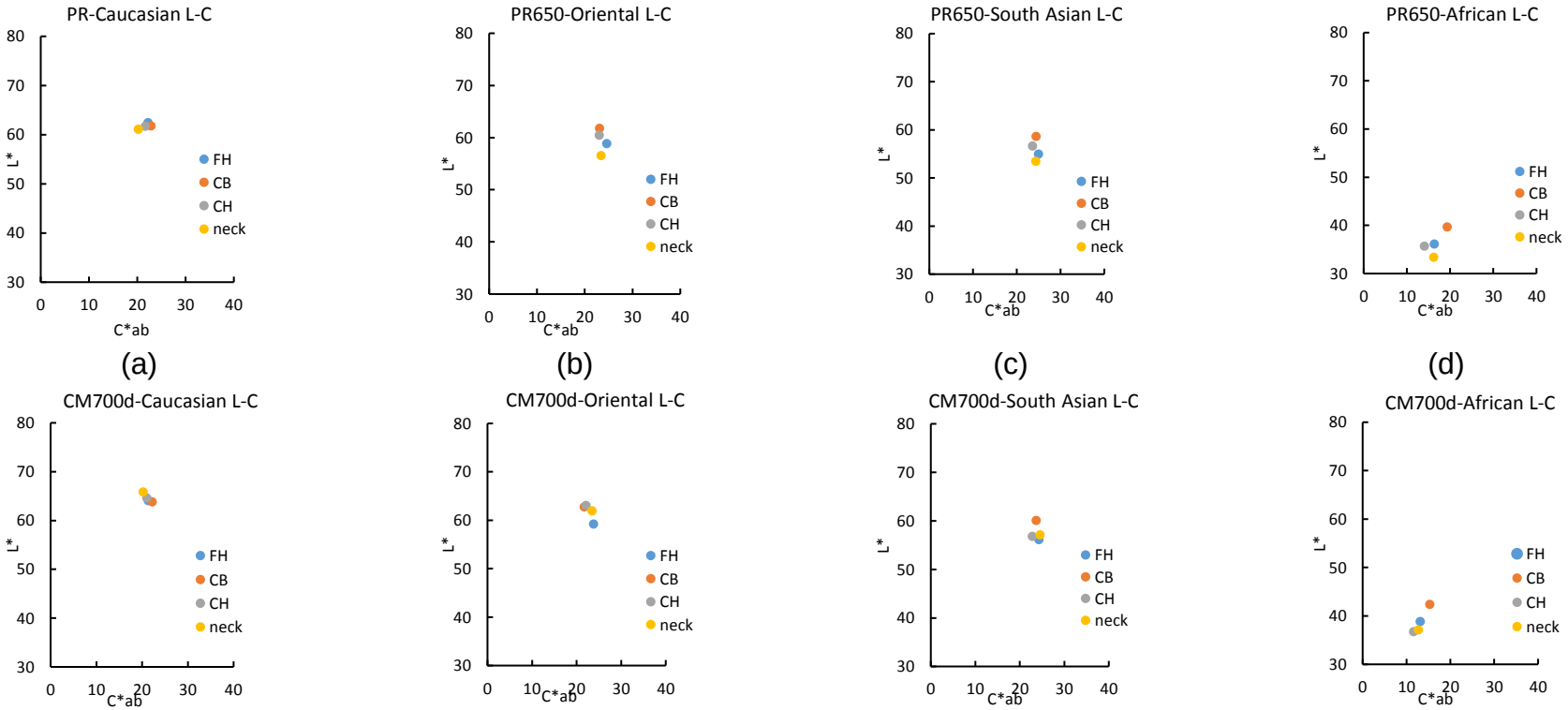


Figure 4.1.2.10 The mean CIELAB values of 4 locations for 4 ethnic groups and plotted in $\mathbf{L^*C^*_{ab}}$plane. measured by the PR650 (PR): (a) the Caucasian group; (b) the Chinese group; (c) the South Asian group; (d) the African group; and measured by the CM700d (CM): (e) the Caucasian group; (f) the Chinese group; (g) the South Asian group; (h) the African group. (FH: forehead; CH: cheek; CB: cheekbone).

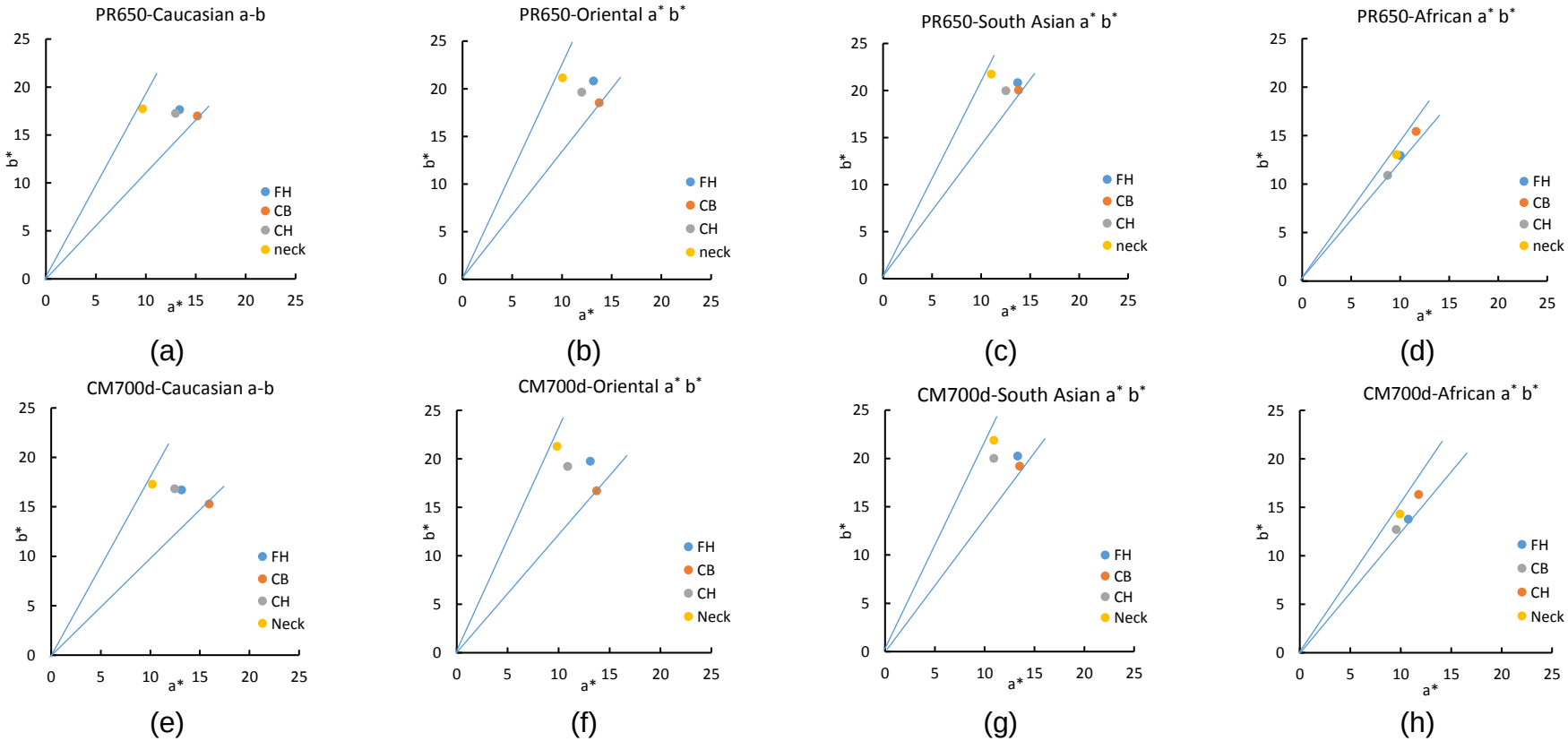


Figure 4.1.2.11 The mean CIELAB values of 4 locations for 4 ethnic groups and plotted in in $\mathbf{a}^*\mathbf{b}^*$plane. measured by the PR650 (PR): (a) the Caucasian group; (b) the Chinese group; (c) the South Asian group; (d) the African group; and measured by the CM700d (CM): (e) the Caucasian group; (f) the Chinese group; (g) the South Asian group; (h) the African group. (FH: forehead; CH: cheek; CB: cheekbone).

The measurement results of the PR650 and CM700d gave similar trends in $L^*C^*_{ab}$ plane, as shown in Figure 4.1.2.10. For the Oriental and South Asian group, as shown in Figure 4.1.2.10 (b)-(c) and (f)-(g), the CB and CH appeared whiter than the FH. For the Caucasian group, these three locations (CB, CH and FH) had similar colour. For the African group, the CB was appeared to be the whitest, followed by FH and CH in that order. The two instruments gave different measurement results for the neck. The measurement results of PR650 of the neck had the lowest lightness amongst the four locations, as shown in Figure 4.1.2.10 (a)-(d). The measurement results of CM700d of the neck had similar lightness value with the CH. It even appeared to be the whitest amongst all four locations of the Caucasian group, as shown in Figure 4.1.2.10 (e). The disagreement of the neck results can be due to the complex curved surface at this location.

The measurement results of the two instruments had similar distributions in a^*b^* plane, as shown in Figure 4.1.2.11. For the Caucasian, Oriental and South Asian groups, the neck had the largest hue angle, followed by CH, FH and CB in that order. This implies that the neck was the yellowest and the CB was the reddest, amongst those four locations. The two instruments had poor agreement for the measurement results of CH and FH. For the CM700d, the FH appeared redder than the CH, except for the Caucasian group (Caucasian FH and CH had similar hue angle). For the PR650, the CH and FH had similar hue angle. As the CM700d is a contact type instrument, this disagreement can be caused by the pressure on or the stretch of the skin at that target location. For the African group, all four locations had limited variation in the hue angle.

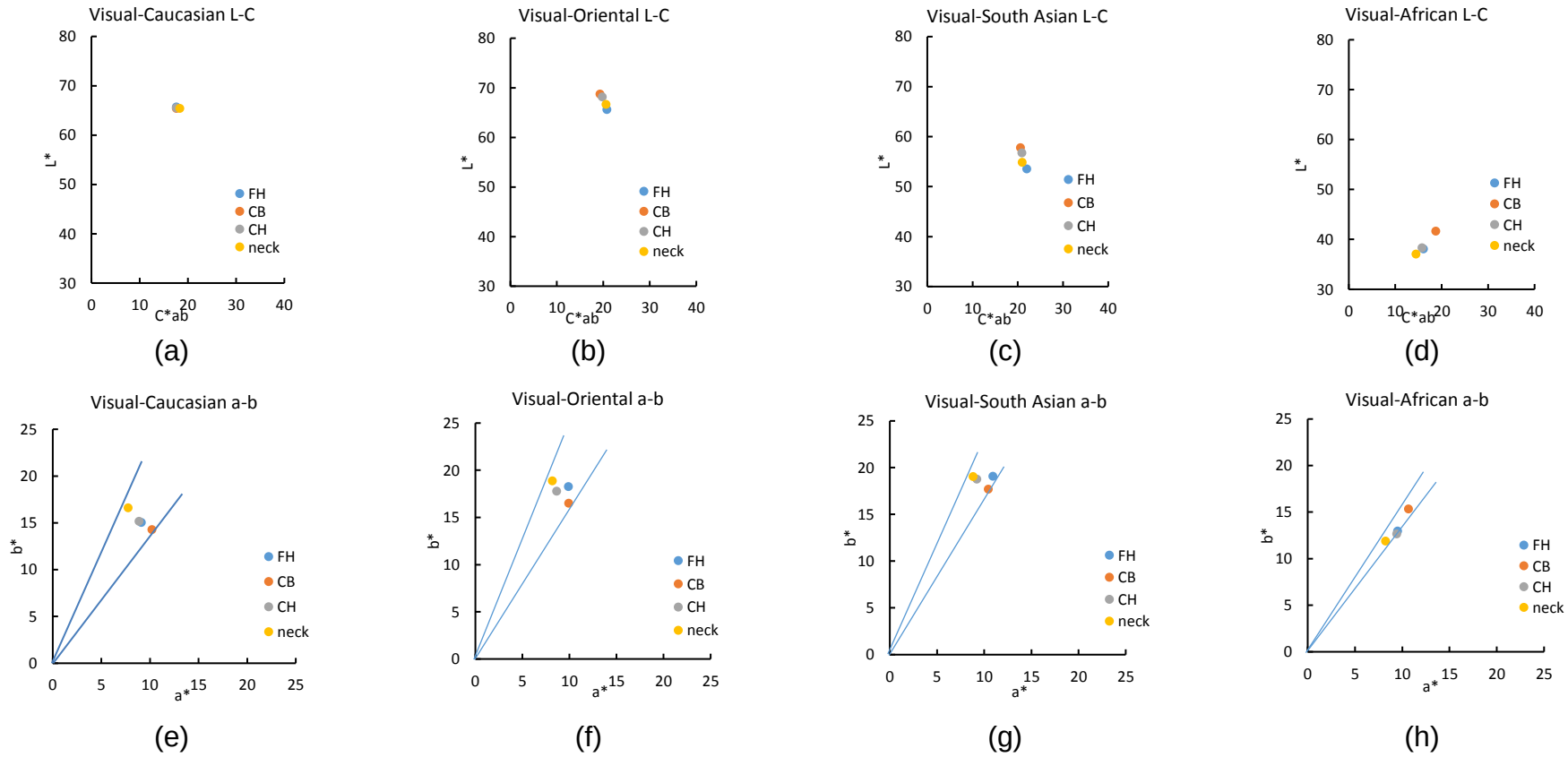


Figure 4.1.2.12 The mean CIELAB values of 4 locations of 4 ethnic groups that were measured by visual assessment (VA) method were plotted in $\mathbf{L^*C^*_{ab}}$ plane: (a) the Caucasian's data in $\mathbf{L^*C^*_{ab}}$plane; (b) the Oriental group; (c) the South Asian group; (d) the African group; and plotted in $\mathbf{a^*b^*}$plane: (e) the Caucasian group; (f) the Oriental group; (g) the South Asian group; (h) the African group. (FH, forehead; CH, cheek; CB, cheekbone).

The visual assessment (VA) is the method that was the closest to human observation. The mean CIELAB of the measurement at each location are shown in Figure 4.1.2.12. Compared to the measurement results of the PR650 and the CM700d, the VA measurement results show similar trends in $L^*C^*_{ab}$ plane. But the measurement results from VA were relatively less spread out than those of the two instruments. The measurement results of the VA method showed limited variation at the four locations for the Caucasian and Oriental groups in $L^*C^*_{ab}$ plane. Also, the FH and CH locations for the Caucasian and African groups appeared very similar to in both planes.

It can be seen from Figure 4.1.2.10 to Figure 4.1.2.12 that, for all three measurement methods, the FH location appeared lower in lightness value and higher in chroma value than the other locations, except for the African group. This can due to the fact that these locations were more directly exposed to sunlight than the others

4.1.2.8.2 Difference between body locations and facial locations

Apart from the four facial locations discussed above, six more locations, including IFA, OFA, nose, chin, FT and BH, were measured using the CM700d. Figure 4.1.2.13 and Figure 4.1.2.14 show the distribution of the mean CIELAB values of each ethnic group in CIELAB a^*b^*plane and CIELAB $L^*C^*_{ab}$ plane for these 10 locations.

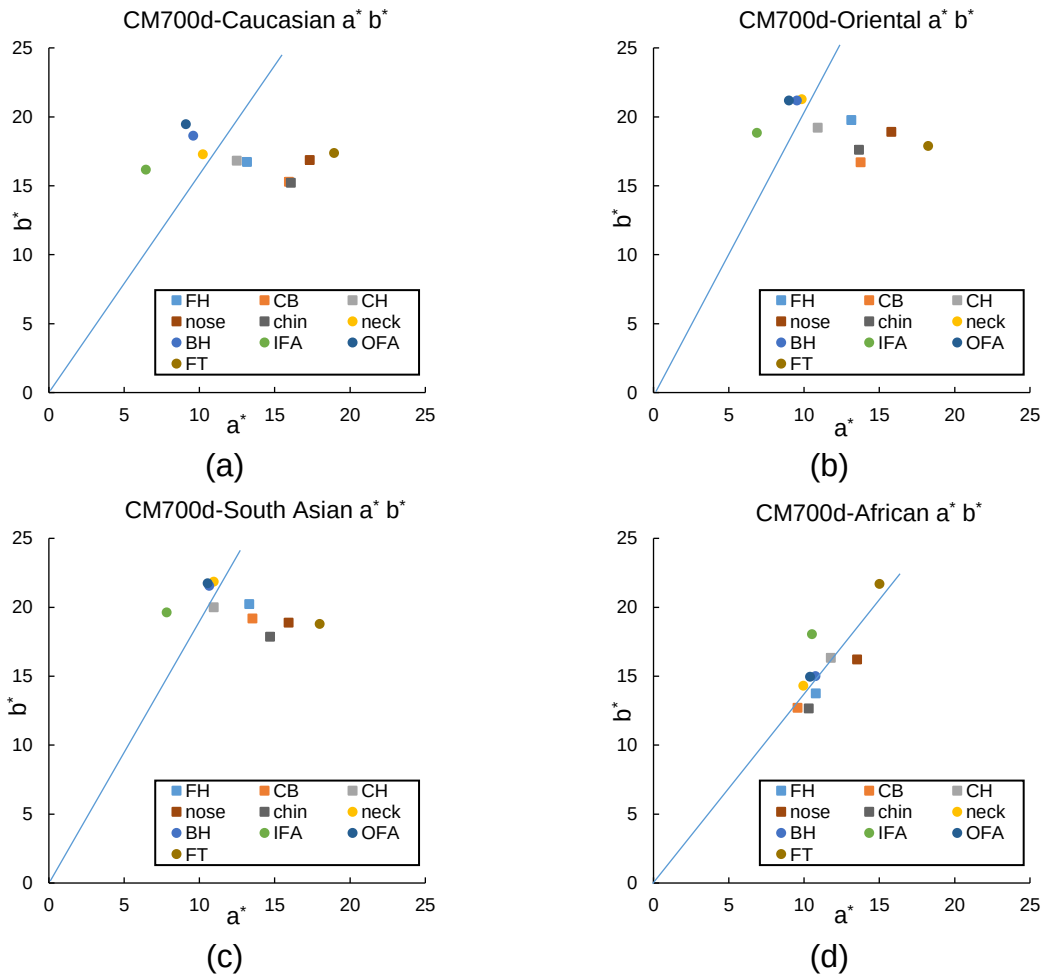


Figure 4.1.2.13 Mean CIELAB values of four ethnic groups at 10 locations that were measured using the CM700d plotted in $\mathbf{a^*b^*}$ plane. The body locations are marked by dots, the facial locations are marked by squares. (a) the Caucasian group; (b) the Oriental group; (c) the South Asian group; (d) the African group.

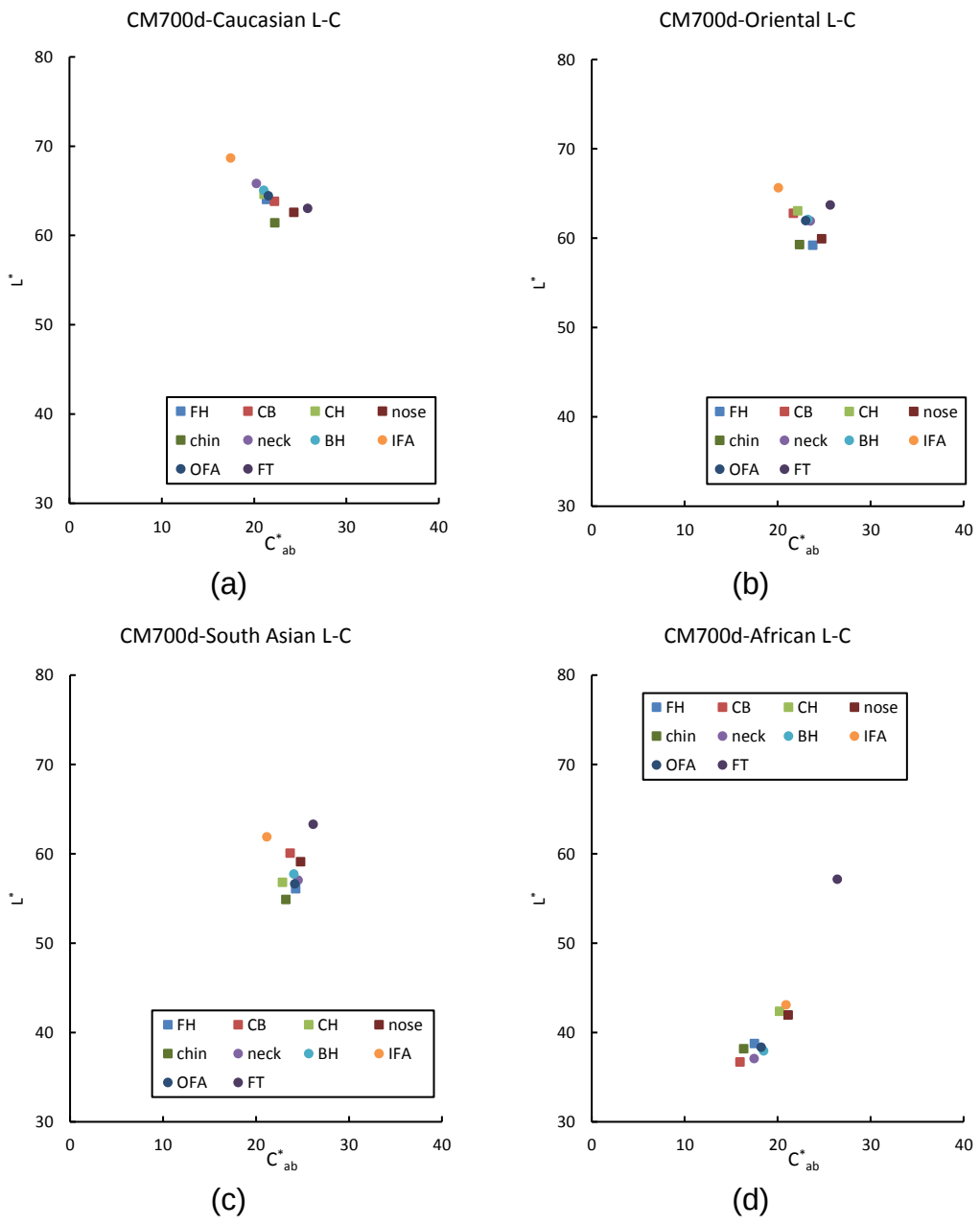


Figure 4.1.2.14 Mean CIELAB values of four ethnic groups at 10 locations that were measured using the CM700d and plotted in $\mathbf{L^*C^*_{ab}}$ plane. The body locations are marked by dots, the facial locations are marked by squares. (a) the Caucasian group; (b) the Oriental group data; (c) the South Asian group; (d) the African group.

In these figures, the facial locations are marked by squares and the body locations are marked by circles. In a^*b^*plane, as shown in Figure 4.1.2.13 (a)-(d), the body locations (exclude the FT) had larger hue angles than the facial locations, except the African group. This implies that the body locations appeared yellower than the facial locations. The similar findings were reported

by Xiao *et al*. (2017) (see Section 2.2.2). For the African group, the ten locations had limited variation in the hue angle, but they had relatively larger variation in chroma. In $\mathrm{L^*C^*_{ab}}$ plane, as shown in Figure 4.1.2.13 (a)-(d), the differences between body locations and facial locations were limited. For the ten locations of the four ethnic groups, apart from FT, the IFA was the whitest. For the Caucasian, Oriental and South Asian, these ten locations had small variation in chroma. Especially for South Asian, the major colour difference between these ten locations was in lightness.

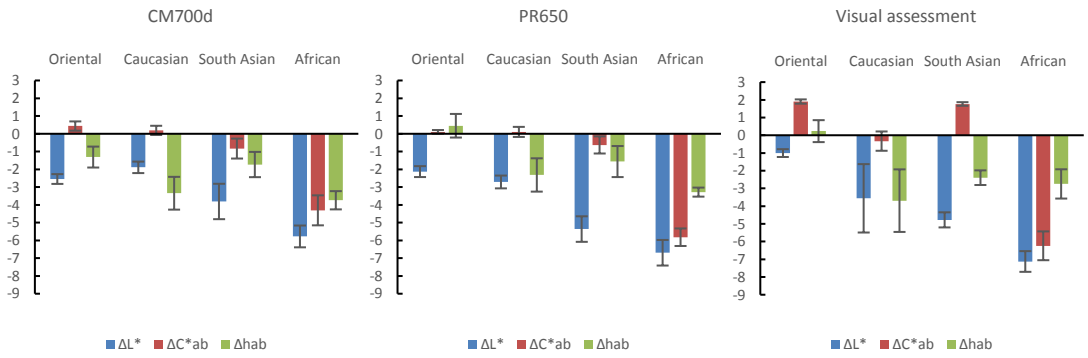
It can be summarised from above that the locations that were exposed to more sun (e.g. the FH, CB, CH and OFA locations) appeared lower in lightness and higher in chroma than the other locations that were less frequently exposed to direct sunlight (such as the IFA and neck), except for the African group. A similar finding was obtained in the last section. These findings validated the skin whiteness trend line which was proposed in Section 4.1.2.5. In that section, the relatively darker skin colour was the one with lower lightness value and higher chroma value which as the same as the finding in this section.

4.1.2.8.3 Difference between genders

This section investigates the difference between different genders. The skin colour difference between the two genders was revealed via the difference in the lightness (ΔL^*), in the chroma (ΔC^*_{ab}) and in the hue angle (Δh_{ab}), as shown in Figure 4.1.2.15. The differences ΔL^*, ΔC^*_{ab} and Δh_{ab} were calculated using the mean CIELAB values of the male minus the mean CIELAB values of the female subjects of the same ethnic group. The error bars represent the standard error of the mean of different measurement locations. Figure 4.1.2.15 (a) and (b) show that the two instruments gave similar results in terms of gender difference. The female subjects had the higher lightness and the larger hue angle compared with the male subjects in all four ethnic groups. This means that the female subjects appeared whiter and yellower than the male subjects. For the PR650 results, both genders of the Caucasian and

Oriental groups had insignificant differences in chroma. But, for the measurement results of CM700d, the Oriental male subjects had a larger value of chroma than the female subjects. For the African and South Asian groups, the male subjects had smaller chroma than female subjects. From the differences in chroma, it can be seen that the skin colour of the Oriental male subjects appeared more saturated than that of the equivalent female subjects, when measured using the CM700d. The skin colour of the South Asian and African male subjects appeared less saturated than the equivalent female subjects. Figure 4.1.2.15 (c) shows the results that were measured using the VA method. As shown in this figure, the skin colour of the Caucasian, South Asian and African male subjects had lower lightness values and smaller hue angles than the equivalent female subjects. The Oriental male subjects also had a lower lightness value than equivalent female subjects, but their values of hue angle did not show significant differences.

It can be concluded that the female subjects had lighter and yellower skin than male subjects, except for the Oriental group results from the CM700d, where the Oriental female subjects had redder skin colour than male subjects. For the difference in chroma, the Caucasian male subjects and female subjects had an insignificant difference in chroma. The Oriental and South Asian male subjects had larger chroma than the female subjects. The African male subjects had smaller chroma than the female subjects. These showed that the Oriental and South Asian male subjects had more saturated skin colour than female subjects, and African male subjects had less saturated skin colour than female subjects.



(a) (b) (c)

Figure 4.1.2.15 The colour shift between two genders in terms of $\Delta \boldsymbol{L}^*$, $\Delta \boldsymbol{C}^*_{\boldsymbol{ab}}$ and $\Delta \boldsymbol{h}_{\boldsymbol{ab}}$. (a) the data from the CM700d; (b) the data from the PR650; (c) the data from the visual assessment.

4.2 Summary of Chapter 4

This chapter has introduced a new skin colour database, the Leeds Liverpool Skin Colour database, that includes the skin colour data that was measured using three measurement methods, a tele-spectroradiometer (the PR650), a spectrophotometer (the CM700d) and a set of colour charts (VA). The facial images that were captured via a facial imaging system also included. Approximately 200 people, including both genders, of four ethnic groups (Caucasian, Oriental, South Asian and African), participated this experiment, and their ten locations (6 facial locations and 4 body locations) were measured. Four facial locations were measured by using three measurement methods. A digital camera was used to capture facial images for all the participants.

The measurement protocols of the two instruments (PR650 and CM700d) was determined by testing the short-term repeatability of two instruments at different settings when measuring human skin colour in vivo (see Section 3.1.2). The short-term (5 continuous measurements at once), medium-term (5 repeat measurement throughout a day) and long-term (the measurement at each experimental day) repeatability of these two instruments and the camera were determined via the MCDM value of the measuring results of the two skin colour patches of MCCC. The MCDM values were listed in Table 4.1.3. Note

that the RGB of skin colour patches from the camera were transformed into CIELAB by using the camera characterisation model which was mentioned in Section 3.2.1.1. The repeatability results showed that the PR650, CM700d and the camera had similar short-term repeatability (about 0.06 ΔE^*_{ab}) satisfactory repeatability. The medium-term and long-term repeatability of the camera system (about 1.70 ΔE^*_{ab}) was less repeatable than the PR650 and CM700d (below 1 ΔE^*_{ab}).

For the VA method, the agreement between three trained observers was determined by calculating the MCDM of the VA results of three trained observers. VA results of 7 subjects were used to determine the agreement between observers and the mean MCDM was 3.05 ΔE^*_{ab}, which agreed with the results that reported by Alfvin and Fairchild (3.03 ΔE^*_{ab}) (1997).

The measurement locations were marked and then measured. Four facial locations were measured by all three tools. The other six locations were only measured via the CM700d. The measurement results were transformed into CIELAB values for analysis.

The spectral reflectance of the four ethnic groups from the PR650 and CM700d were compared in terms of measurement instruments and ethnic groups. Comparing the measurement results of two instruments, the spectral reflectance from two instruments agreed well. The absorption of the CM700d at the long wavelength range (about 650 nm to 700 nm) was slightly more than that of the PR650. Comparing the spectral reflectance of four ethnic groups, the Caucasian group reflected back higher energy, followed by Oriental, South Asian and African in the order.

The distributions of the measurement data showed that different measuring tools gave similar measurement data distributions in CIELAB $L^*C^*_{ab}$ and CIELAB a^*b^* planes. The distribution of the skin colour in $L^*C^*_{ab}$ plane was similar to the shape of a banana. The Caucasian group had higher lightness, followed by Oriental, South Asian and African groups in that order.

The range of the hue angle of different measurement methods and the camera were similar. The range of the hue angle of measurement results of VA methods was slightly smaller than the others. This can due to the range of the hue angle of the skin colour chart that used is small.

Based on the trend of the skin colour distribution in $L^*C^*_{ab}$ plane, formulae to calculate the skin whiteness and skin blackness were proposed. These formulae were developed by adopting the Berns' formulae (2014) for calculating *Vividness* and *Depth.* The skin whiteness and blackness formula were compared with the Individual Typology Angle, ITA˚, which has been used in the cosmetics industry to describe the skin whiteness via L^* and b^*. A positive linear correlation was found between these indices. This validated the proposed formulae.

Comparing the measurement results from different methods, the results from the PR650 and the CM700d were similar. The CM700d results had a slightly higher lightness and lower chroma. The VA results were higher in lightness and appeared yellower than those from the two instruments.

Comparing different locations, the skin colour data of the Oriental and South Asian groups showed that the CB had the highest lightness value, followed by CH and FH in that order. These three locations of the Caucasian group had similar lightness and chroma. The African group data showed that the CB also had the highest lightness, followed by FH and CH in that order. For the neck, different measurement methods gave different results: the PR650 measurement results for the neck had the lowest lightness among the four locations; the CM700d measurement results for the neck had similar lightness values with the CH. This disagreement might be due to the complex curved surface of this location. Comparing the body locations and facial locations, the body locations appeared yellower than the facial locations, except for the African group. The different locations of the African group had limited variation in the hue angle, but had a relatively large variation in chroma.

The locations that were more frequently exposed to the sun had a lower lightness and higher chroma. This finding also validated the skin whiteness trend line which is proposed in the present study.

Comparing the skin colour of different genders, the female skin colours had higher lightness than that of the male subjects. The skin colour of the Oriental male subjects appeared more saturated than that of the equivalent female subjects when the CM700d was the measurement instrument. Also, the skin colour of the South Asian and African male subjects appeared less saturated than that of the equivalent female subjects.

Chapter 5 Facial skin colour and perceived facial impression

Experiment 1, described in Chapter 4, accumulated skin colour and facial images of different facial and body locations of different ethnic groups, to build the LLSC database. Based on this database, two psychophysical experiments were designed to investigate the influence of skin colour on the perceived facial impression attributes. The first experiment (Experiment 2.1) aimed to investigate the relationships between the attributes when judging the images with different skin colours. Twenty-three attributes were examined and a smaller number of attributes were selected to represent these twenty-three attributes. Based on the first experiment, the second experiment (Experiment 2.2) was to investigate the impact of skin colour on the selected attributes. The influence of gender and ethnic group were also studied.

The display of the skin colour of the facial image in the psychophysical experiment was crucial to this investigation. It is important that the displayed colour be both repeatable and accurate. Before the main experiment was carried out, the repeatability and accuracy of the displayed skin colour were investigated (see Section 3.2). The range of the skin colours that were investigated in this research was determined from the LLSC database. In the present study, the experimental results were analysed in terms of skin whiteness and hue angle. The skin whiteness can be calculated from CIELAB coordinates (see Section 4.1.2.5). All the experimental results were first transformed into z-scores. The z-scores were then plotted in line charts and bubble charts, according to different skin whiteness values and hue angles.

5.1 Experimental preparation

5.1.1 Attributes for describing facial impression

Twenty-three scales were selected and investigated in the present study. They are listed in Table 5.1.1. The attributes marked in yellow, red and blue are the attributes that related to the facial impression, personality and skin appearance, respectively.

Table 5.1.1 Selected attributes presented in both English and (simplified) Chinese.

Perceived Impression		Skin Appearance	
English	Chinese (simplified)	English	Chinese (simplified)
Dull-Lively	沉闷的-活泼的	Pale-Pink	苍白的-红润的
Feminine-Masculine	女性化的-男性化的	Tan-Fair	黝黑的-白皙的
Unnatural-Natural	不自然的-自然的	Opaque-Translucent	不通透的-通透的
Stupid-Intelligent	笨的-聪明的	Dry-Moist	干燥的-水嫩的
Babylike-Mature	不成熟的-成熟的	Matt-Glossy	无光泽-有光泽
Dislikeable-Likable	不讨人喜欢的-讨人喜欢的	Rough-Smooth	粗糙-光滑
Serious-Funny	严肃的-幽默的	Blemished-Clear	有瑕疵的-无瑕疵的
Unhealthy-Healthy	不健康的-健康的		
Unattractive-Attractive	吸引人的-不吸引人的		
Aged-Youth	年老的-年轻的		
Fussy-Easygoing	挑剔的-温和的		
Tensed-Relaxed	焦虑的-从容的		
Ordinary-Imaginative	平庸的-想象力丰富的		
Loner-Cooperative	独来独往的-乐于合作的		
Passive-Active	被动的-主动的		
Autistic-Sociable	孤僻的-合群的		

In this table, ten terms are included in the impression section. These terms were studied by the previous researchers (see Section 2.2.3) to investigate the impact of the facial features (e.g. eye size, face shape, skin colour, etc.) on the impressions, except *Dull-Lively* and *serious-Funny*. The *Dull-Lively* and *Serious and funny* were the terms appeared in the cosmetic product

advertisements, but they were not being studied in the past. Six terms in the personality section were selected from the big five model adjective list (McCrae and Costa Jr, 1999). The terms in these two sections that mentioned above were named perceived impression attributes in the present study. Seven terms in the skin appearance section were mostly selected from a study of Wang *et al*. (2016). The translucency and the moisture were from the studies of Tsumura, *et al*. (2008) and of Jiang and Kaplan (2011). The attributes were translated into Chinese (Simplified) by using the Oxford Chinese dictionary. The attributes in Chinese were used in the experiments, as only the Chinese observers took part these two experiments.

5.1.2 Images Preparation

As mentioned in Section 2.2.3, the morphed face image was frequently used in face related study. But this kind of image lost the skin texture information due to the morph process. This can lead to the quality of the image reduced, especially at the contrast. To have the observing situation close to the real life, the facial image of a subject was selected to be the stimuli images. To examine the impact of skin colour on the perceived facial impressions, the stimuli images that only differed in skin colour were used. The stimuli images were achieved through image processing.

The image processing workflow is given in Figure 5.1.2.1 and two stages are included. The first stage is skin colour segmentation (marked in the blue frame). In this stage, the facial skin area was extracted from the original image. An example of the extracted area is marked as image A, as shown in Figure 5.1.2.1. The second stage is to apply a new skin colour on the extracted skin area (marked in the red frame). In this stage, the texture of the extracted area was first taken out and stored. Then, the new skin colour was applied onto the facial skin area and the stored texture was also added back in. The facial area with the new skin colour is shown in image B in Figure 5.1.2.1. Finally, the new facial skin area was restored back to the image and the image processing was completed.

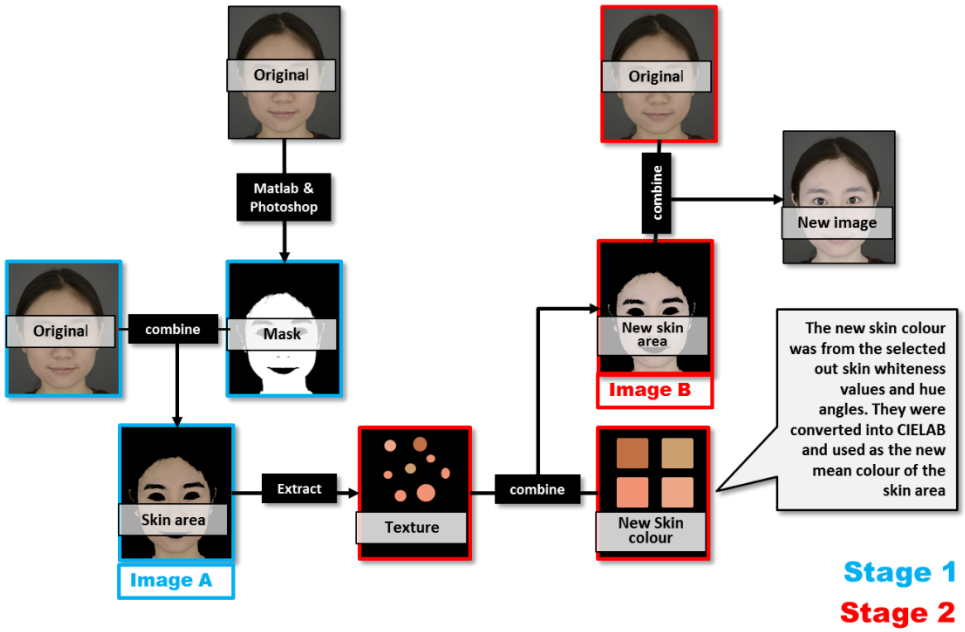


Figure 5.1.2.1 Image processing workflow to generate the various facial images.

Four original images, including two ethnic groups (Oriental and Caucasian) with both genders, were chosen. They represent the typical good appearance of each group from the LLSC database for generating stimuli images. *MATLAB® R2011a* and *Adobe® Photoshop CS6* were used to carry out the image segmentation and extraction. These two stages were explained in Sections 5.1.2.1 and 5.1.2.2, respectively. The structure of the MATLAB code is shown in Figure B1 in Appendix B. The new images were contracted and stored with the monitor dependent RGB. The GOG model was used to transform from RGB values to XYZ tristimulus values (see Section 3.2.2.5).

Stage 1 Facial skin area segmentation

Stage one aimed to extract the facial skin area from the images. The segmentation algorithm used was a threshold based segmentation method (Jones and Rehg, 2002; Al-Tairi *et al*., 2014). The skin area was identified through the range of the RGB values of the skin colour. This range was determined by using the digital facial images in the LLSC database. The RGB values of the skin colour were extracted from four facial locations (FH, CB,

CH, and neck) of each image. Then the maximum and minimum of these RGB values were used as the range of the skin colour. The maximum and minimum RGB values of the Oriental and the Caucasian group's skin colour are listed in Table B1 (Appendix B). Based on the skin colour range, a mask was generated, as shown in Figure 5.1.2.1. This mask had the skin area and the non-skin area marked in white and black, respectively. The non-skin area included some of the facial features, such as the eye, eyebrows and lips. The sclera of the eye was hard to identify by using this segmentation method. So, *Adobe® Photoshop CS6* (selection and paint function) was used to mark this area in black. Finally, the skin area in the image was extracted using the mask.

Stage 2 Rendering new skin colour

The second stage was to render the new skin colour on the facial skin area which was extracted in the first stage. Before the rendering procedure, the texture of the skin was extracted and stored. A statistical image analysis method was adopted to use in the present study to extract the skin texture (Young and Liu, 1986). In the original method, the texture of a neutral colour digital image is considered as the variation of each pixel in luminance, against the mean luminance of all the pixels. So, for the colour image, the texture can be considered as the variation of the L^*, $a^*$and b^* at each pixel, against the mean L^*, $a^*$and b^* value of all the pixels. This can be expressed by using Equation 5.1.

$$\begin{bmatrix} \Delta L^*_{i,j} \\ \Delta a^*_{i,j} \\ \Delta b^*_{i,j} \end{bmatrix} = \begin{bmatrix} L^*_{i,j} \\ a^*_{i,j} \\ b^*_{i,j} \end{bmatrix} - \begin{bmatrix} L^*_{mean} \\ a^*_{mean} \\ b^*_{mean} \end{bmatrix} \qquad \textbf{Equation 5.1}$$

where $L^*_{i,j}$, $a^*_{i,j}$ and $b^*_{i,j}$are the CIELAB values of the pixel (i,j). L^*_{mean}, a^*_{mean} and b^*_{mean} are the mean CIELAB of the all the pixels. $\Delta L^*_{i,j}$, $\Delta a^*_{i,j}$, $\Delta b^*_{i,j}$ are the extracted texture information of the pixel (i,j).

The new skin was rendered by adding the original texture to the selected skin colour (see Equation 5.2). The skin colour selection will be further explained in the next section.

$$\begin{bmatrix} L^*_{new\ i,j} \\ a^*_{new\ i,j} \\ b^*_{new\ i,j} \end{bmatrix} = \begin{bmatrix} \Delta L^*_{i,j} \\ \Delta a^*_{i,j} \\ \Delta b^*_{i,j} \end{bmatrix} - \begin{bmatrix} L^*_{new} \\ a^*_{new} \\ b^*_{new} \end{bmatrix} \qquad \textbf{Equation 5.2}$$

where $\Delta L^*_{new\ i,j}$, $\Delta a^*_{new\ i,j}$ and $\Delta b^*_{new\ i,j}$ are the CIELAB values of the new skin pixel (i,j). $\Delta L^*_{i,j}$, $\Delta a^*_{i,j}$ and $\Delta b^*_{i,j}$ are the skin texture that gained through Equation 5.1. L^*_{new}, a^*_{new} and b^*_{new} are the selected new skin colour.

Then the skin colour area of the original image was replaced by the new skin which had the same texture but a different colour. Finally, an image with new skin colour was generated.

5.1.3 The new skin colour selection

The new skin colours were selected based on two rules. The first rule is that the selected new skin colours should cover the skin colour gamut. The skin colour gamut could be determined through the LLSC database. The skin colour data in the LLSC database were arranged in terms of CIELAB $\mathrm{L^*a^*b^*}$ or $\mathrm{L^*C^*_{ab}h_{ab}}$ coordinates. It is complicated to select the skin colours that can cover the skin colour gamut evenly in a three-dimensional colour space. In Section 4.1.2.5, a formula to estimate skin whiteness was proposed. This formula transforms the lightness and chroma of the skin colour into a skin whiteness index. This means that the lightness and chroma of the skin colour can be transformed into the skin whiteness value. So, the skin colour can be described in a two-dimensional colour space (skin whiteness and hue angle) instead of using the three-dimensional colour space (e.g. lightness, chroma and hue angle). The complexity of the new skin colour selection was thus

reduced. The second rule is that the selected new skin colour should also include some skin colours outside the skin colour gamut. This ensured the whole skin colour gamut was covered.

Based on these conditions, the CIELAB values of the skin colour of Caucasian and Oriental subjects in the LLSC database were transformed into values of skin whiteness and the hue angle. Then, the skin colours that covered the skin colour range in the LLSC were selected evenly. Figure 5.1.3.1 shows the plots of the Caucasian and Oriental skin colour in the skin whiteness and the hue angle plane (marked in grey). Note that the (hue angle, skin whiteness) of original images were not plotted in this figure, so only 25 dots included in each sub-figure. The selected skin colours are marked in red in these figures. The selected new skin colours are also plotted in CIELAB $L^*C^*_{ab}$ and CIE L^*h_{ab} planes, as shown in Figures B2 to B5 in Appendix B.

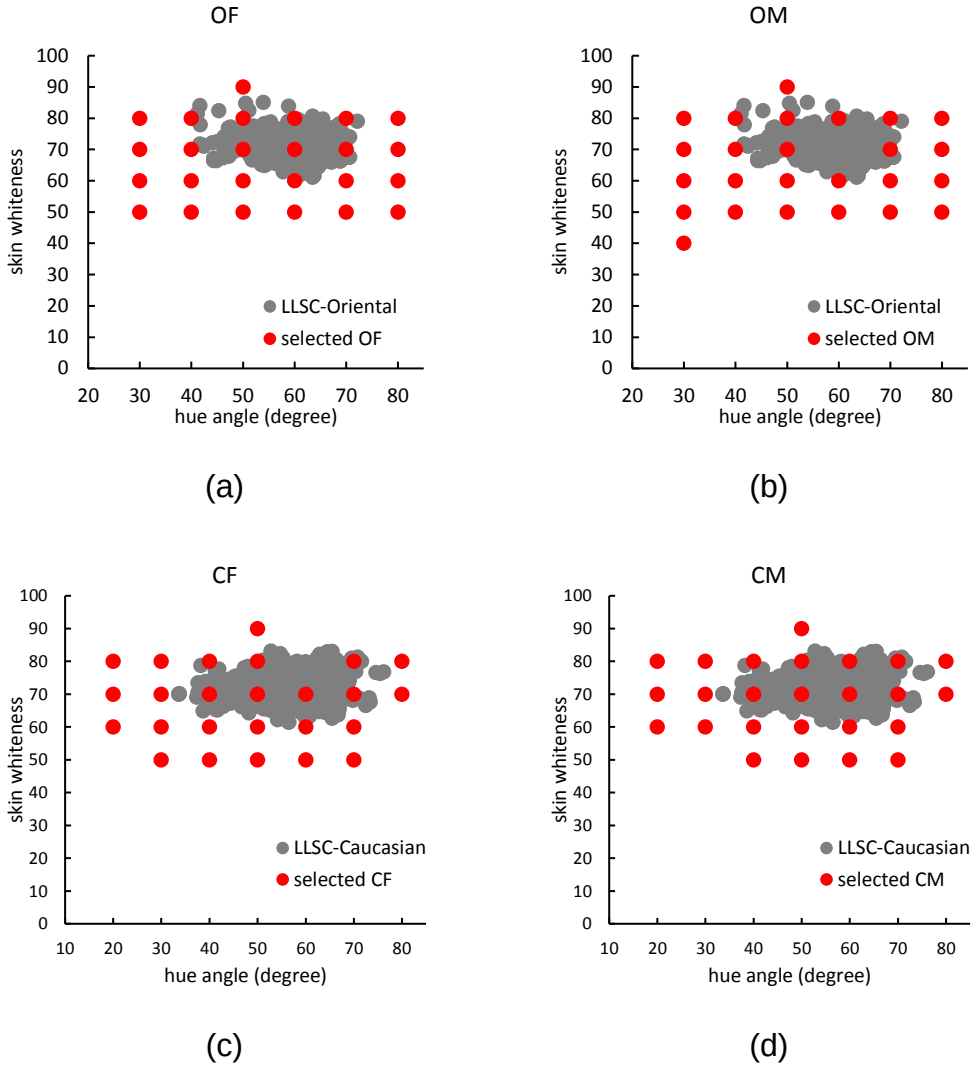


Figure 5.1.3.1 The skin colour gamut of the Oriental (a)-(b) and Caucasian (c)-(d) (marked in grey) and the selected new skin colour (marked in red): (a) the selected new skin colour for Oriental female images; (b) the selected new skin colour for Oriental male images; (c) the selected new skin colour for Caucasian female images; (d) the selected new skin colour for Caucasian male images.

As mentioned above, the selected skin colours were described in terms of skin whiteness and hue angle. But the image processing method used in the present study was based on CIELAB L^*, a^* and b^* coordinates. Therefore, these selected skin colours defined in whiteness and hue angle need to be first transformed into lightness, chroma and hue angle, and then to CIELAB (D65/2). This transformation needs two equations that are related to the skin whiteness, lightness and chroma. Apart from the skin whiteness formula, an

equation that described the relationship between the L^* and C^*_{ab} was needed. Here, a skin whiteness trend line, named personalised skin whiteness trend line, was proposed to use. This trend line is based on the skin colour a subject. It was generated by using the *Skin Whiteness coefficient and* the mean skin colour (CIELAB) of this subject. The mean skin colour was the average of the CIELAB values of the four facial locations. In the present study, the measurement results from PR650 were used. This trend line described the gradual variation of the skin whiteness of each subject. The formula of the personalised skin whiteness trend line is given in Equation 5.3.

$$L^* = \alpha C^*_{ab} + \beta \qquad \textbf{Equation 5.3}$$

where $\alpha = \frac{L^*_{Sub} - \beta}{C^*_{ab_{sub}}}$ and L^*_{Sub} and $C^*_{ab_{sub}}$ are the mean lightness and chroma value of the skin colour of a subject which were calculated from the average CIELAB values of the four facial locations that were measured by using the PR650. L^*_{Sub} and $C^*_{ab_{sub}}$ values are given in Table 5.1.2. β is the coefficient optimised from the LLSC database which are listed in Table 4.1.11. The personalised skin whiteness trend lines are also listed in Table 5.1.2.

Table 5.1.2 The $\mathbf{L^*_{Sub}}$ and $\mathbf{C^*_{ab_{sub}}}$ of the four subjects (CM, CF, OF and OM) and their personalised skin whiteness trend line.

	L^*_{Sub}	$C^*_{ab_{sub}}$	**Personalised skin whiteness trend line**
CM	66.83	17.99	y = -0.8009x + 80.7
CF	57.18	24.85	y = -0.7706x + 80.7
OM	63.74	21.38	y = -0.7571x + 76.0
OF	66.83	17.99	y = -0.5732x + 76.0

It can be seen from this table that the personalised skin whiteness trend lines have good agreement with the skin whiteness trend lines (Caucasian and Oriental) that were shown in Figure 4.1.2.6. Here, the personalised trend lines

were used for more accurate transformation from skin whiteness to the lightness and chroma,

Then, the L^* and C^*_{ab} values corresponding to a given whiteness value can be obtained by solving the two linear equations (Equation 5.3 and Equation 4.1). Equation 4.1 is rewritten here as $W_s = 100 - \sqrt{(\beta - L^*)^2 + (C^*_{ab})^2}$. The solution of these two equations can be expressed as the equations in the Equation 5.4.

$$L^* = \alpha \frac{100 - W_s}{\sqrt{\alpha^2 + 1}} + \beta$$

Equation 5.4

$$C^*_{ab} = \frac{100 - W_s}{\sqrt{\alpha^2 + 1}}$$

where W is the known skin whiteness value which required to transform into L^* and C^*_{ab}.

The above L^* and C^*_{ab} values together with the hue angle were then transformed to L^*, $a^*$and b^* values which were used to generate the rendered images (see Equation 5.2).

The selected skin colour included skin colours that are out of the skin colour range and therefore could lead to the facial images appearing unreal. This could lead to the negative judgement to these ‘unreal facial images’. Then these judgement may not be useful for building the model. However, the reason for selecting skin colours that are out of the range of the skin gamut of the LLSC database is due to the supposition thatsubjects that were included in the database may not fully cover the skin colour of this ethnicity. Even though the selected in such cases the skin colour may lead to the image appearing unreal, this process can ensure the range of the ‘real’ skin colour can be included.

5.2 Experiment 2

The skin colour data of four different ethnic groups by using different measurement methods were accumulated in Experiment 1. Based on these skin colour data, Experiment 2 was aimed to investigate the influence of the skin colour on the perceived impressions and skin appearance. Here, the Chinese observers were recruited and studied. Two psychophysical experiments, named Experiments 2.1 and Experiment 2.2, were conducted. The aim of Experiment 2.1 was to determine the relationship between selected attributes when observing images only different in skin colour. In this experiment, the number of the attributes could be reduced. The aim of Experiment 2.2 was to further investigate the impact of the skin colour, gender and ethnic group on the facial image judgement. In this experiment, the influence of the skin whiteness and hue angle on the perceived judgement of facial impression attributes was investigated. The impact of the gender and ethnic group of the subject in the stimuli image and the gender of the observer on the judgements were also studied.

5.2.1 Experiment 2.1

5.2.1.1 Experimental design of Experiment 2.1

In this experiment, two original images, including a Chinese female image and a Chinese male image, were used. Each original image was developed into 26 images that only differed in skin colour. The image processing was explained in Section 5.1.2. In total, 52 images were examined in this experiment. Four images were randomly selected from the 52 images to be used to determine the stability of the observer's judgement. These four images appeared twice in the experiment. Thus, for each observer, 56 images were examined with 23 attributes. Section 5.1.1 explained the selection of the attributes.

Experiment 2.1 used a *Yes or No* method. Each of the 23 bi-polar attribute pairs (as shown in Table 5.1.1), such as attractive and unattractive, were shown below the image in a random order. The observer were asked to select the term that could more appropriately describe the image. Ten Chinese observers, including 5 Chinese females and 5 Chinese males, age from 23 to 28, volunteered to participate in this experiment. The experimental interface and experimental procedure are shown in Figure 5.2.1.1.

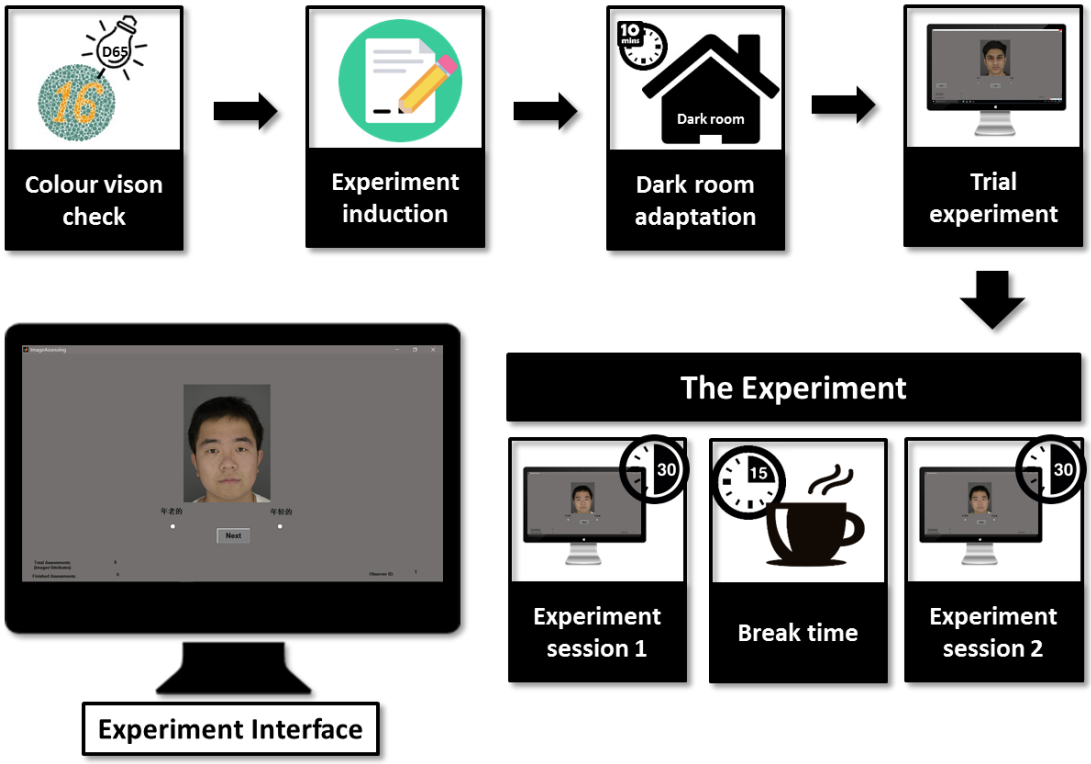


Figure 5.2.1.1 The procedure and interface of Experiment 2.1.

As shown in Figure 5.2.1.1, the colour vision of ten observers were checked by using the *Ishihara's Test For Colour Deficiency (*24 plates) under the simulated D65 illumination first. Then the observers who passed the test stayed in the dark environment for 10 minutes for adaptation. After adaptation, a training test, including 10 images and 23 attributes, was carried out. This training test was aimed to assist the observer to become familiar with the experimental interface. The images used in the training were different from the images that were examined in the experiment. The experiment was

divided into two observation sessions and each session lasted about 30 minutes. Fifteen mins rest time was allowed between the sessions. In this experiment, 1288 judgements were made by each observer ($28\ \text{images} \times 2\ \text{genders} \times 23\ \text{attributes}$).

5.2.1.2 Results and discussion

5.2.1.2.1 Intra-observer and inter-observer variation

The intra-observer and inter-observer variation were evaluated by using a *wrong decision* method (see Section 2.6.2 for details). Figure 5.2.1.2 shows the mean *wrong decisions* from the judgements of 10 Chinese observers with 23 questions. The intra-observer variations and inter-observer variation were marked in blue and orange, respectively. The mean *wrong decision* of 23 question was shown at the right-hand side of the figure. The attributes at the right-hand side of the hyphen were used to represent the attribute pair.

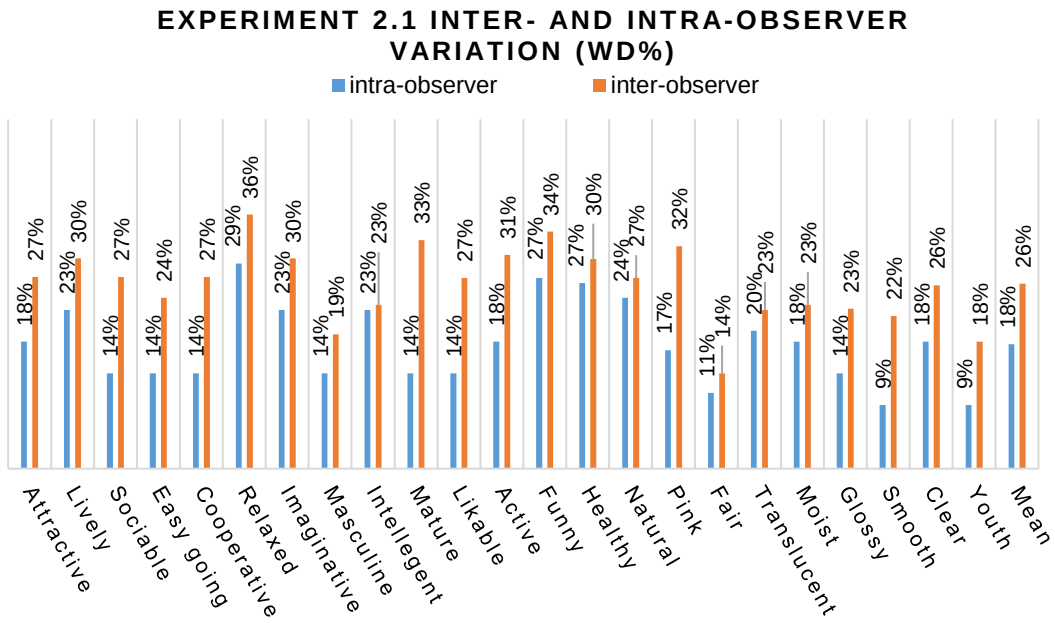


Figure 5.2.1.2 Mean inter-observer and intra-observer variation (wrong decision: WD%).

Figure 5.2.1.2 shows that the mean inter-observer wrong decisions were larger than the intra-observer variation. For the inter-observer variations, the Chinese observers had the largest disagreement at judging the *Relaxed* attribute. The Chinese observers had the best agreement at judging the *Fairness* attribute. For the intra-observer variation, the judgement of the *Relaxed* attribute was poorer in consistency than the others. The judgement of the *Youth* and *Smooth* attributes were the most consistent of all.

5.2.1.2.2 Principal component analysis

Table 5.2.1 shows the 3 components extracted from 23 attributes (listed in Table 5.1.1) by using PCA analysis. The attributes of Component 1, 2 and 3 are marked in red, green and blue, respectively. Figure 5.2.1.3 is the biplot of these three components. In Figure 5.2.1.3, the attributes on the right-side of the hyphen were used to represent the attribute pair and the attributes from Component 1, 2 and 3 were marked in red, green and blue dots.

Table 5.2.1 Rotated Component Matrix of 23 attributes.

	Component		
	1	2	3
Rotation Sums of Squared Loadings % of Variance	38.281	34.576	9.730
Unhealthy-Healthy	0.911	0.082	-0.258
Autistic-Sociable	0.883	-0.033	0.097
Unnatural-Natural	0.871	0.231	-0.290
Loner-Cooperative	0.842	0.082	0.207
Fussy-Easygoing	0.828	-0.020	-0.220
Dislikeable-Likable	0.825	0.374	0.019
Serious-Funny	0.786	-0.209	0.159
Dull-Lively	0.775	-0.099	0.313
Passive-Active	0.759	0.036	0.320
Unattractive-Attractive	0.748	0.560	0.018
Tensed-Relaxed	0.732	0.485	-0.078
Ordinary-Imaginative	0.683	0.316	0.439
Babylike-Mature	0.646	-0.388	0.300
Opaque-Translucent	0.010	0.967	0.061
Tan-Fair	-0.083	0.965	0.059
Aged-Youth	0.107	0.941	-0.063
Matt-Glossy	0.130	0.910	0.032
Dry-Moist	0.144	0.907	0.157
Rough-Smooth	0.046	0.872	0.416
Blemished-Clear	0.066	0.741	0.614
Pale-Pink	0.560	-0.705	-0.144
Stupid-Intelligent	0.459	0.702	0.363
Feminine-Masculine	-0.054	-0.436	-0.855

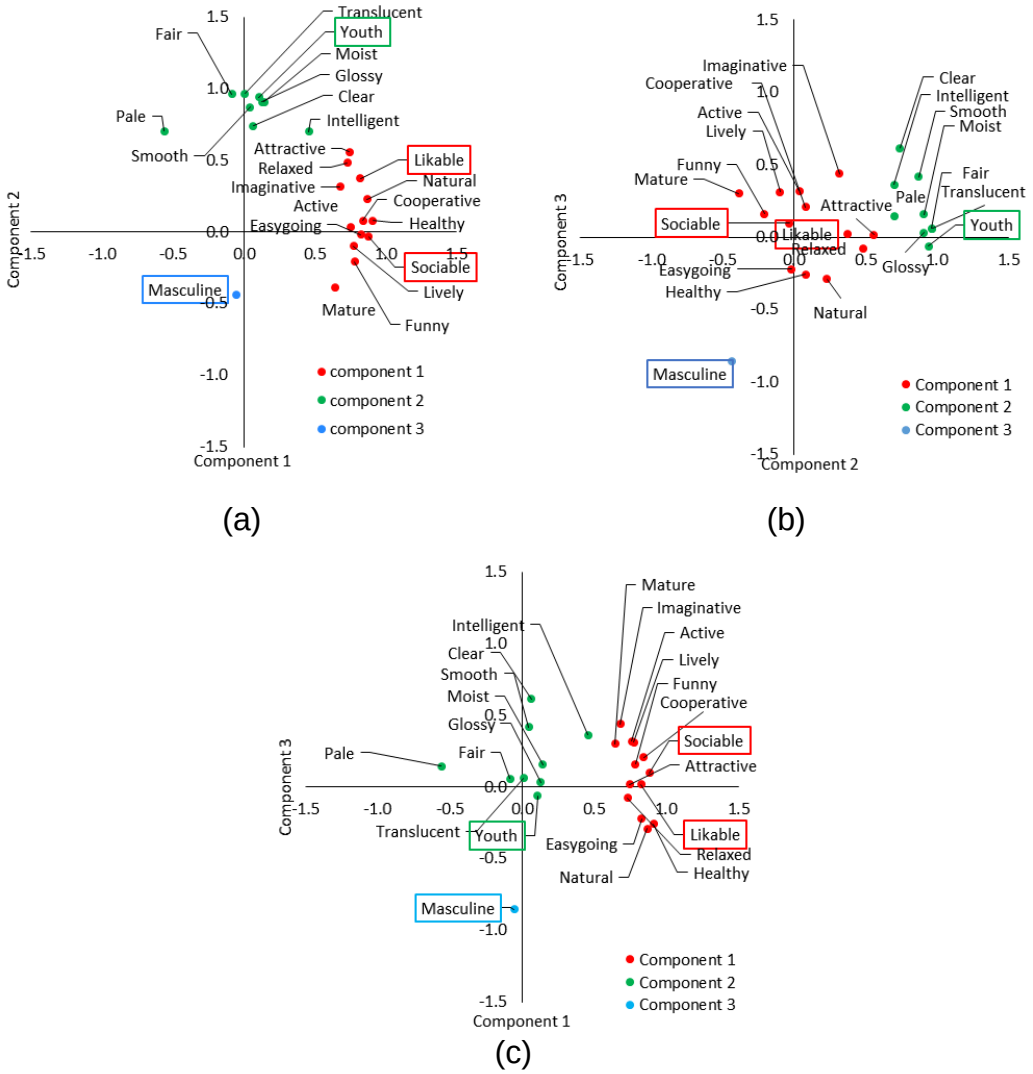


Figure 5.2.1.3 The biplot of the PCA analysis results (Table 5.2.1): (a) component 1 vs 2; (b) component 2 vs 3; (c) component 2 vs 3.

In Table 5.2.1, twenty-three attributes were divided into three groups that belong to Component 1, Component 2 and Component 3, respectively. Component 1 included 13 attributes from the perceived impression group. Component 2 included all seven attributes of the skin appearance group and two attributes from perceived impression group (*Aged-Youth* and *Stupid-Intelligent*). Component 3 included only the *Feminine-Masculine* attribute. It can be seen from these components that the attributes from the same group were strongly correlated, except the *Aged-Youth* and *Stupid-Intelligent* attributes which were correlated to the skin appearance attributes. In Figure

5.2.1.3, the attributes in Component 1 had a large spread. This implies that there is a potential for more components to be extracted from the attributes in Component 1. Thus, the PCA analysis was carried out on the attributes in Component 1. The results of this analysis are shown in Table 5.2.2. The attributes in Component 1a and Component 1b are marked in yellow and orange respectively.

Table 5.2.2. Rotated Component Matrix of the attributes in Component 1.

	Component	
	1a	1b
Rotation Sums of Squared Loadings % of Variance	40.376	33.654
Unattractive-Attractive	0.903	0.220
Tensed-Relaxed	0.886	0.198
Dislikeable-Likable	0.827	0.386
Unnatural-Natural	0.826	0.378
Unhealthy-Healthy	0.732	0.507
Ordinary-Imaginative	0.582	0.486
Fussy-Easygoing	0.573	0.556
Babylike-Mature	0.073	0.827
Serious-Funny	0.318	0.778
Dull-Lively	0.368	0.754
Autistic-Sociable	0.527	0.719
Loner-Cooperative	0.569	0.652
Passive-Active	0.484	0.627

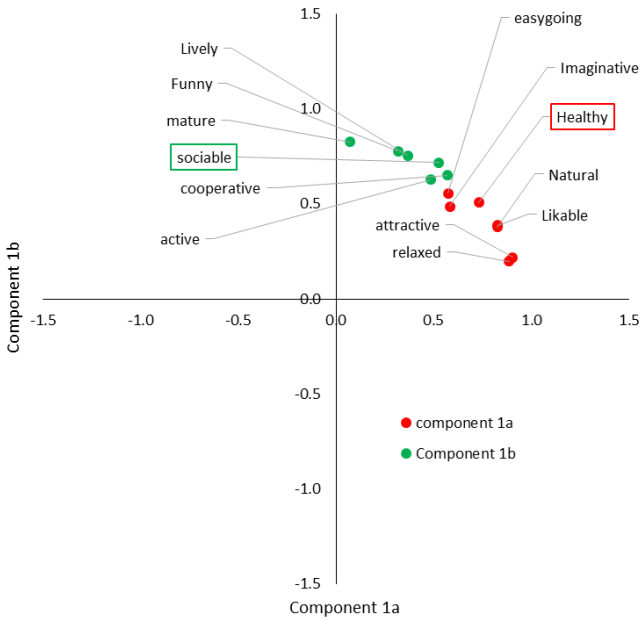


Figure 5.2.1.4 The biplot of the PCA analysis results (Table 5.2.2) plot of the two components: Component 1a attributes (red dot); Component 1b attributes (blue dot).

Two components (Component 1a and Component 1b) were extracted from the attributes in Component 1, as shown in Table 5.2.2. The Component 1a included 7 attributes and Component 1b included 6 attributes. The biplot of these two components is illustrated in Figure 5.2.1.4. In this figure, some of the attributes had very similar loadings, such as *Nature* and *Likeable*, *Relaxed* and *Attractive*, and *Lively* and *Funny*. This indicates that these attributes had the stronger correlation with each other.

The PCA analysis was also carried out on the attributes included in Component 2. The analysis results showed that only one component was included, as shown in Table C2 in Appendix C.

Table 5.2.3 lists the correlation coefficients between 23 attributes, which were used to further explore the internal relationship between these attributes. The attributes from Component 1a, 1b, 2 and 3 were marked in yellow, orange, green and blue, respectively. An absolute r value above 0.6 was considered as strong correlation or better agreement, and they are marked in red.

Table 5.2.3 The correlation coefficients between 23 attributes (the $|r| \geq 0.60$ were marked in red).

	Unattractive-Attractive	Tensed-Relaxed	Dislikeable-Likable	Unnatural-Natural	Unhealthy-Healthy	Ordinary-Imaginative	Fussy-Easygoing	Babylike-Mature	Serious-Funny	Dull-Lively	Autistic-Sociable	Loner-Cooperative	Passive-Active	Opaque-Translucent	Tan-Fair	Agedness-Youth	Matt-Glossy	Dry-Moist	Rough-Smooth	Blemished-Clear	Pale-Pink	Stupid-Intelligent	Feminine-Masculine
Unattractive-Attractive	–	0.79	0.84	0.78	0.73	0.66	0.54	0.28	0.44	0.48	0.65	0.66	0.65	0.53	0.45	0.61	0.64	0.60	0.53	0.48	0.05	0.72	−0.32
Tensed-Relaxed	0.79	–	0.78	0.77	0.71	0.64	0.61	0.27	0.46	0.52	0.60	0.61	0.48	0.47	0.37	0.52	0.46	0.50	0.42	0.38	0.01	0.65	−0.20
Dislikeable-Likable	0.84	0.78	–	0.77	0.73	0.67	0.68	0.40	0.57	0.62	0.68	0.75	0.60	0.37	0.29	0.43	0.47	0.42	0.35	0.36	0.18	0.60	−0.25
Unnatural-Natural	0.78	0.77	0.77	–	0.90	0.59	0.72	0.42	0.58	0.51	0.71	0.66	0.60	0.20	0.13	0.32	0.31	0.29	0.12	0.04	0.35	0.46	0.09
Unhealthy-Healthy	0.73	0.71	0.73	0.90	–	0.54	0.81	0.51	0.63	0.58	0.78	0.67	0.60	0.06	−0.01	0.15	0.19	0.20	0.00	−0.03	0.51	0.40	0.10
Ordinary-Imaginative	0.66	0.64	0.67	0.59	0.54	–	0.44	0.48	0.52	0.64	0.61	0.58	0.61	0.35	0.27	0.35	0.35	0.48	0.47	0.52	0.10	0.70	−0.51
Fussy-Easygoing	0.54	0.61	0.68	0.72	0.81	0.44	–	0.54	0.55	0.56	0.70	0.66	0.53	0.00	−0.10	0.05	0.02	0.08	−0.04	−0.04	0.43	0.29	0.16
Babylike-Mature	0.28	0.27	0.40	0.42	0.51	0.48	0.54	–	0.54	0.49	0.65	0.49	0.43	−0.33	−0.40	−0.32	−0.31	−0.21	−0.21	−0.05	0.51	0.20	−0.13
Serious-Funny	0.44	0.46	0.57	0.58	0.63	0.52	0.55	0.54	–	0.69	0.70	0.67	0.59	−0.19	−0.22	−0.12	−0.06	−0.01	−0.08	−0.02	0.58	0.25	−0.08
Dull-Lively	0.48	0.52	0.62	0.51	0.58	0.64	0.56	0.49	0.69	–	0.64	0.73	0.70	−0.04	−0.12	0.02	0.01	0.08	0.11	0.12	0.50	0.35	−0.22
Autistic-Sociable	0.65	0.60	0.68	0.71	0.78	0.61	0.70	0.65	0.70	0.64	–	0.75	0.65	−0.03	−0.09	0.07	0.10	0.08	0.05	0.09	0.47	0.47	−0.12
Loner-Cooperative	0.66	0.61	0.75	0.66	0.67	0.58	0.66	0.49	0.67	0.73	0.75	–	0.73	0.10	0.04	0.16	0.18	0.20	0.21	0.22	0.38	0.54	−0.26
Passive-Active	0.65	0.48	0.60	0.60	0.60	0.61	0.53	0.43	0.59	0.70	0.65	0.73	–	0.04	−0.04	0.10	0.22	0.19	0.24	0.29	0.42	0.43	−0.30
Opaque-Translucent	0.53	0.47	0.37	0.20	0.06	0.35	0.00	−0.33	−0.19	−0.04	−0.03	0.10	0.04	–	0.94	0.90	0.86	0.91	0.88	0.75	−0.66	0.69	−0.46
Tan-Fair	0.45	0.37	0.29	0.13	−0.01	0.27	−0.10	−0.40	−0.22	−0.12	−0.09	0.04	−0.04	0.94	–	0.89	0.86	0.89	0.86	0.73	−0.74	0.67	−0.45
Agedness-Youth	0.61	0.52	0.43	0.32	0.15	0.35	0.05	−0.32	−0.12	0.02	0.07	0.16	0.10	0.90	0.89	–	0.86	0.82	0.80	0.64	−0.59	0.72	−0.34
Matt-Glossy	0.64	0.46	0.47	0.31	0.19	0.35	0.02	−0.31	−0.06	0.01	0.10	0.18	0.22	0.86	0.86	0.86	–	0.85	0.79	0.72	−0.48	0.66	−0.45
Dry-Moist	0.60	0.50	0.42	0.29	0.20	0.48	0.08	−0.21	−0.01	0.08	0.08	0.20	0.19	0.91	0.89	0.82	0.85	–	0.86	0.77	−0.53	0.73	−0.54
Rough-Smooth	0.53	0.42	0.35	0.12	0.00	0.47	−0.04	−0.21	−0.08	0.11	0.05	0.21	0.24	0.88	0.86	0.80	0.79	0.86	–	0.91	−0.66	0.76	−0.71
Blemished-Clear	0.48	0.38	0.36	0.04	−0.03	0.52	−0.04	−0.05	−0.02	0.12	0.09	0.22	0.29	0.75	0.73	0.64	0.72	0.77	0.91	–	−0.58	0.75	−0.86
Pale-Pink	0.05	0.01	0.18	0.35	0.51	0.10	0.43	0.51	0.58	0.50	0.47	0.38	0.42	−0.66	−0.74	−0.60	−0.48	−0.53	−0.66	−0.58	–	−0.33	−0.39
Stupid-Intelligent	0.72	0.65	0.60	0.46	0.40	0.70	0.29	0.20	0.25	0.35	0.47	0.54	0.43	0.69	0.67	0.72	0.66	0.73	0.76	0.75	−0.33	–	−0.63
Feminine-Masculine	−0.32	−0.20	−0.25	0.09	0.10	−0.51	0.16	−0.13	−0.08	−0.22	−0.12	−0.26	−0.30	−0.46	−0.45	−0.34	−0.45	−0.54	−0.71	−0.86	−0.39	−0.63	–

It can be concluded from this table that the attributes of the same components had a stronger correlation with each other. For Component 1a, *Dislikeable-Likable* and *Tensed-Relaxed* have stronger correlation with other attributes within this component. For Component 1b, *Autistic-Sociable* has stronger correlation with other attributes within this component. The *Autistic-Sociable* also had a stronger correlation with all the attributes within Component 1a and 1b. For Component 2, *Tan-Fair*, *Aged-Youth* and *Rough-Smooth* have stronger correlation to the other attributes. The *Unattractiveness-Attractiveness* attribute was the attribute that also had a stronger correlation with some of the attributes in Component 2. The *Unattractiveness-Attractiveness* attribute was found to have a stronger correlation with the skin appearance attributes, *Matt-Glossy* and *Dry-Moist*. For the three attributes that related to the skin spatial appearance (*Matt-Glossy*, *Dry-Moist* and *Rough-Smooth*), it can be seen from this table that the *Matt-Glossy* attribute had a stronger correlation with all the skin appearance attributes, except *Pale-Pink*; the *Dry-Moist* attribute is associated with *Tan-Fair*, *Opaque-Translucent* and *Rough-Smooth*; the *Rough-Smooth* attribute is closely related with *Blemishes-Clear*.

Four attributes were selected to represent these components based on two rules: firstly, this attribute needs to have a strong correlation with other attributes within the same component; secondly, this attribute needs to be easily understood by Chinese observers. Based on these two rules, the *Autistic-Sociable, Dislikeable-Likable and Feminine-Masculine* can be selected out to represent the impression attributes of Component 1a, 1b and 3. For the Component 2, the *Tan-Fair*, *Aged-Youth* and *Rough-Smooth* were all found suitable to be selected to represent the Component 2. Here, the *Aged-Youth* was selected, as this impression attribute was widely studied in previous research work. The *Unhealthy-Healthy* attribute was also selected to investigate in the present study, as a healthy is an attribute of wide interest for traditional Chinese medicine. The attribute with positive meaning in the attribute pairs was selected, except the *Feminine-Masculine* attribute which had no positive or negative meaning. Based on the translation and easy understanding by the Chinese observer, the *Feminine* attribute was chosen to

use. Finally, the *Sociable, Healthy, Youth, Feminine* and *Likable* attributes were selected to use in Experiment 2.2.

Note here, the *Youth* attribute is used here to express all the appearance word-pairs, i.e. a more youthful skin colour would appear fairer, smoother, clearer, and more pinkish, translucent, moist and glossy.

5.2.1.3 Summary of Experiment 2.1

Experiment 2.1 investigated the underlying pattern of the attributes by using PCA analysis. Through this experiment, the size of the attributes can be reduced. The inter-observer and intra-observer variation (WD%) were 18% and 26%, respectively. By using PCA analysis, four components were extracted and four attributes were selected to represent them. These four attributes were *Sociable*, *Feminine*, *Youth* and *Likeable*. The *Healthy* attribute was also selected for investigation in Experiment 2.2, as it has great interest in traditional Chinese medicine.

5.2.2 Experiment 2.2

5.2.2.1 Experimental design of Experiment 2.2

In this experiment, four original images, including two ethnic groups (Oriental and Caucasian) with both genders, were examined. Each original image was processed into 26 further images via image processing (Section 5.1.2). The images of the Chinese female and the Chinese male were the same images that were used in Experiment 2.1. The *Paired Comparison* method was used in this experiment. The comparison was only carried out between the stimuli images that were generated from the same original images. In total, there are 1300 pairs of images ($325\ pairs\ of\ images \times 4\ original\ images = 1300$) were examined in this experiment. Twenty-four Chinese observers, including 12 Chinese females and 12 Chinese males, participated in this experiment. Figure 5.2.2.1 shows the experimental procedure and the experimental interface.

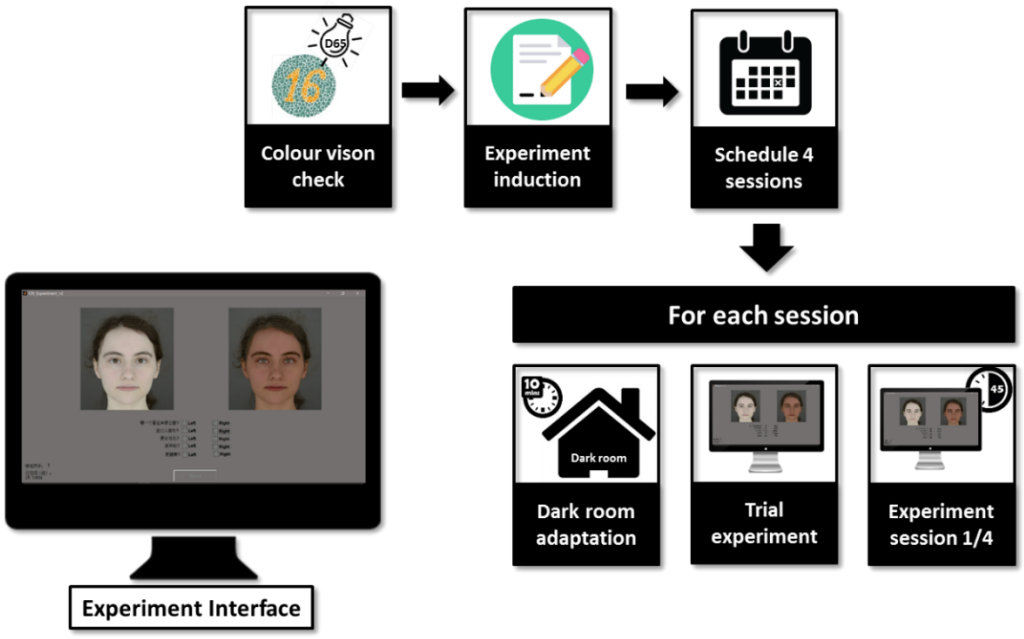


Figure 5.2.2.1 The procedure and interface of Experiment 2.2.

As shown in the Figure 5.2.2.1, the observers judged the same pair of stimuli images based on 5 questions on each trial. These five questions were:

For the facial images above, which one looks more Sociable? Which one looks more Feminine? Which one looks Younger? Which one looks more Likeable? Which one looks Healthier?

An investigation of the difference between answering one question and answering five questions at the same time was carried out before this experiment. It was demonstrated that the results between these two experimental designs had moderate agreement (see Appendix C for detailed information). This means that a different questioning method has limited influence on the experimental results. Thus, in the present study, the experiment was designed to answer 5 questions at the same time, as the overall size of the experiment was smaller.

Experiment 2.2 was divided into four sessions, each session including approximately 340 pairs of images. About 45 minutes were needed to complete each session. Similar to the Experiment 2.1, repeat image pairs were included in each session to determine the stability of the observer. In this experiment, six pairs of stimuli images in each session were randomly selected and used to investigate the stability. For each observer, four sessions were completed within one week. As for Experiment 2.1, the colour vision of the observers was first checked using the *Ishihara's Test for Colour Deficiency* (24 plates) under the D65 illumination. Also, the observers who passed the test stayed in the dark environment for 10 minutes for adaptation. After adaptation, a training test, including 10 pairs of stimuli images, was carried out. This training trial was carried out before every session. The images used in the training were different from the stimuli images that were examined in this experiment.

5.2.2.2 Results and discussion

5.2.2.2.1 Inter-observer and intra-observer variation

The inter-observer variation and intra-variation were determined via the *wrong decision* method. The intra-observer variation was accessed by the repeat judgements of 6 pairs of stimuli images in each session. The intra-observer variation of each observer was determined via the mean *wrong decision* of all four sessions. The mean *wrong decision* of all 24 observers for each attribute was calculated and is shown in Figure 5.2.2.2. The inter-observer variation was calculated to investigate the variation of the judgement between observers. The mean *wrong decision* of 24 observers of each attribute was used to determine this variation, as shown in Figure 5.2.2.2. The intra-observer and inter-observer variation are marked in blue and orange in the figure, respectively.

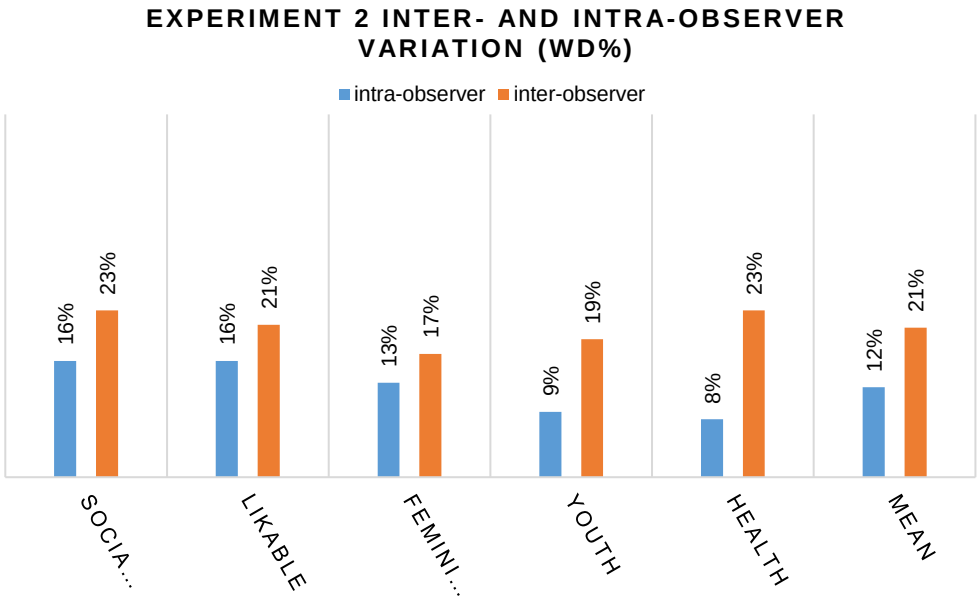


Figure 5.2.2.2 The bar chart of the inter-observer and intra-observer variation (wrong decision: *WD%*).

From Figure 5.2.2.2, it can be seen that, for the intra-observe variation, the Chinese observer gave a more consistent response to the skin appearance

attributes than the perceived impression attributes. The average wrong decision of the skin appearance attributes and the perceived impression attributes was 15% and 8.5%, respectively.

For the inter-observer variation, the mean *wrong decision* of these five attributes was 21%. Similar to the Experiment 2.1, the inter-observer variation was larger than the intra-observer variation, as shown in Figure 5.2.2.2. In these 5 attributes, the Chinese observers showed good agreement at judging the *Feminine* attribute. They gave less agreed response on the *Sociable* and *Health* attributes.

5.2.2.2.2 Agreement between female and male observers

The correlation coefficient (*r*) was used to determine the agreement between the judgements of the female and male Chinese observers. The agreement was investigated in terms of different skin whiteness, different hue angle, different images and different attributes. Table 5.2.4 and Table 5.2.5 list the *r* values in terms of different skin whiteness and different hue angle, respectively. The last row of each table lists the mean *r* of the five attributes at each image. The last column of each table lists the mean *r* of the four images. The average *r* of all the *r* values in the table is underlined in the table. The *r* value below this average *r* value is marked in yellow. The judgement results were also plotted in a scatter diagram, as shown in Figures D9 to D12 (Appendix D).

Table 5.2.4 The *r* of the judgement between two genders with images varied in whiteness.

Images / Attributes	OF	OM	CF	CM	Mean
Sociable	0.82	1.00	0.96	0.98	**0.94**
Likable	0.76	0.95	0.97	0.99	**0.92**
Feminine	0.84	0.97	0.99	0.95	**0.94**
Youth	0.95	0.85	0.99	1.00	**0.95**
Health	0.80	0.99	0.96	0.98	**0.93**
Mean	**0.83**	**0.95**	**0.97**	**0.98**	**0.93**

Table 5.2.4 lists the *r* value of the judgements made by the female and the male Chinese observers of the images with different skin whiteness. From the mean *r* value of the five attributes, it can be seen that the female and male Chinese observers had better agreement at judging the CF and CM images, but they had less agreeable judgements of the OF images. The mean *r* value of judging OF images was 0.83. For the OF images, the two gender groups had the least agreeable response for the *Likeable* attribute (*r*=0.76). They gave the most agreeable judgement for the *Youth* attribute (*r*=0.95). For the OM images, the male and female Chinese observers were least agreeable (*r*=0.85) at judging the *Youth* attribute. The judgement of the *Sociable* attribute had the best agreement (*r*=1.00). The two gender groups had good agreement in judging the CF and CM images at five attributes: the *r* values were greater than 0.95. From the mean *r* value of four images, it can be found that the agreement between the female and male Chinese observers for five attributes were similar.

Table 5.2.5 The *r* of the judgement between two genders with images varied in hue angle**.**

Images / Attributes	OF	OM	CF	CM	Mean
Sociable	0.98	0.93	1.00	0.98	**0.97**
Likable	0.99	0.94	0.98	0.89	**0.95**
Feminine	0.96	0.99	0.48	0.68	**0.78**
Youth	0.97	0.98	0.77	0.36	**0.77**
Health	0.96	0.83	0.98	0.93	**0.92**
Mean	**0.97**	**0.93**	**0.84**	**0.77**	**0.88**

Table 5.2.5 lists the *r* values of the judgements made by the female and male Chinese observers of the images that varied in hue angle. In the last row, the female and male Chinese observers had the poor agreement at judging the Caucasian images compared to the Chinese images. For the OF image, the *r* values of all five attributes were greater than 0.95. For the OM images, the agreement for the *Health* was the poorest (*r*=0.83). The agreement for the *Feminine* attribute was the best (*r*=0.99). For the CF images, the poorest agreement was at judging the *Feminine* attribute (*r*=0.48). The best agreement appeared when judging the *Sociable* attribute (*r*=0.97). For the CM image, the agreement for the *Youth* attribute was the poorest (*r*=0.36). The judgement of the *Sociable* attribute had the best agreement (*r*=0.98). From the last column, it can be found that the two gender groups had a poorer agreement when judging the Youth attribute (*r*=0.77) and the Feminine attribute (*r*=0.78). The judgement of *Sociable* attribute had the best agreement (*r*=0.97).

The average *r* values of the two gender groups when judging varied skin whiteness and hue angle were 0.93 and 0.88, respectively. This showed that the agreement at judging skin whiteness was slightly better than when judging the hue angle.

From the above analysis, it can be found that the agreement of the Chinese observers with different gender could be influenced by the images and the attributes. The best correlation can be as high as r=1.00. The poorest agreement can be as low as r=0.36. The judgement results with an r value below the average were plotted in line charts, as shown in Figure 5.2.2.3 and Figure 5.2.2.4. The judgement result of the male and female Chinese observers is marked in blue and orange, respectively. The skin whiteness and the hue angle that are between the two dashed lines were the skin whiteness or the hue angle within the LLSC database. The other judgement results were also plotted in the line charts in terms of the attributes, images, skin whiteness and hue angle, as shown in Appendix D Figures D1 to D8.

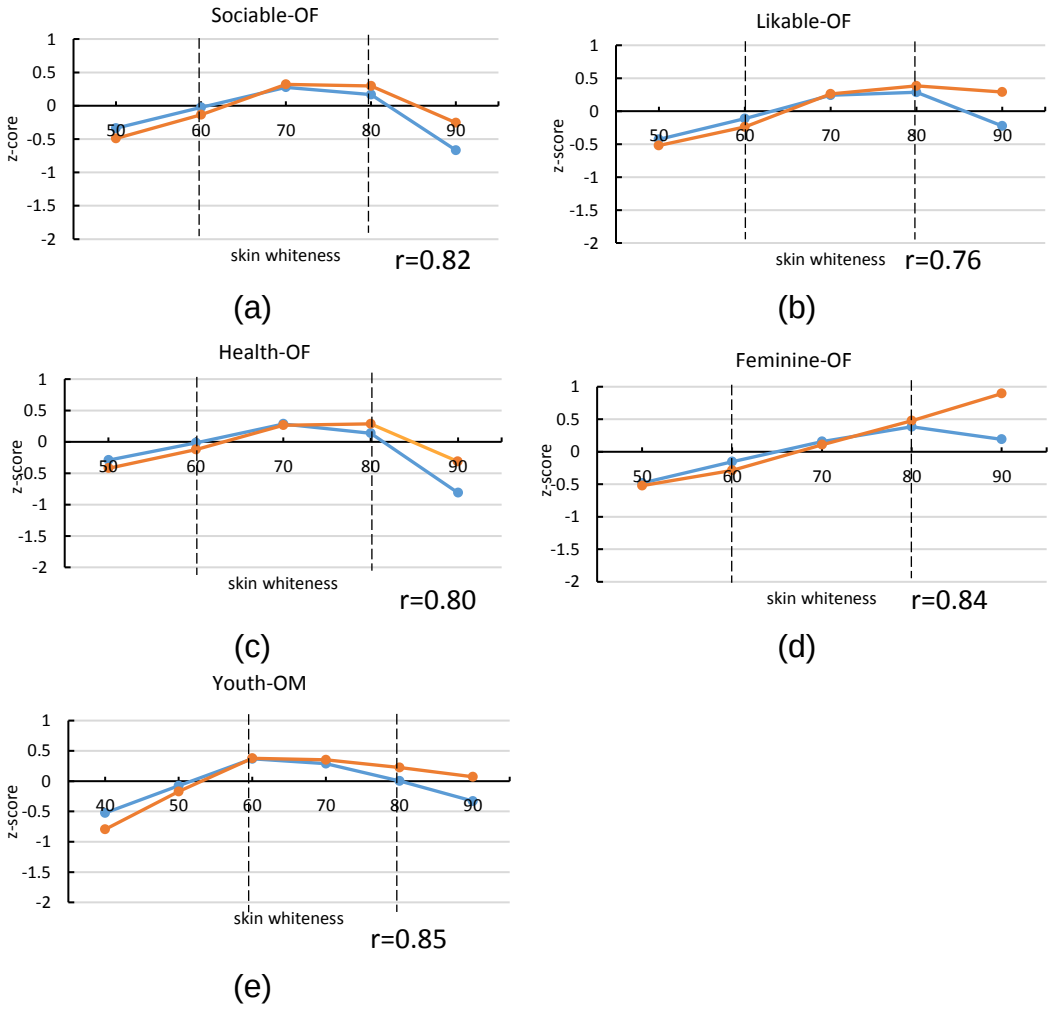


Figure 5.2.2.3 The poor agreement at judging the images with different skin whiteness: (a) judge the sociable of the Chinese female image;(b) judge the Likeable of the Chinese female image;(c) judge the Health of the Chinese female image; (d) judge the Feminine of the Chinese female image; (e) judge the Youth of the Chinese male image.

From Figure 5.2.2.3, it can be found that the poor agreements between female and male Chinese observers were appeared mostly at judging the images with high skin whiteness. The male Chinese observers gave a stronger response when judging the *Sociable* and *Health* attributes of the OF image with skin whiteness equal to 90, as shown in Figure 5.2.2.3 (a) and (c). The male Chinese observers also gave a stronger response when judging the *Likeable* attribute of the Chinese male with skin whiteness equal to 90, as shown in Figure 5.2.2.4 (f). The female Chinese observers gave a stronger response when judging the *Feminine* attribute of the OF images with skin whiteness

equal to 90. The male and female Chinese observers hold opposite opinions when judging the *Likeable* attribute of the OF image with a skin whiteness equal to 90. The female Chinese observers thought the OF images appeared likeable but the male Chinese observers thought these OF images appeared dislikeable. They also held opposite opinions when judging the *Youth* attribute of the OM images with skin whiteness equal to 90. The female Chinese observers thought the OM images appeared youthful but the male Chinese observers thought the OM images appeared aged.

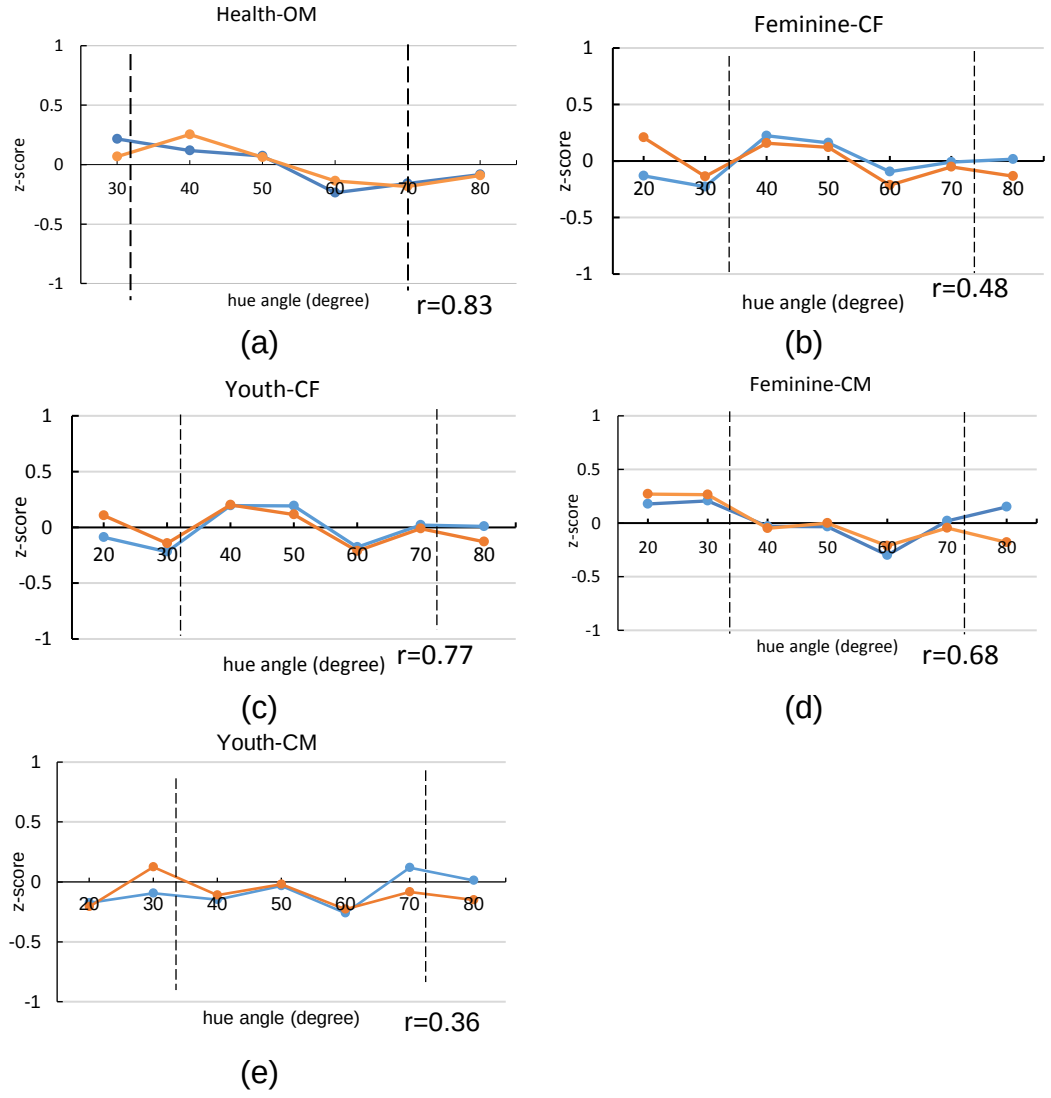


Figure 5.2.2.4 The poor agreement at judging the images with different hue angles of skin colour: (a) judge the Health of the Chinese male image;(b) judge the Feminine of the Caucasian female image;(c) judge the Youth of the Caucasian female image; (d) judge the Feminine of the Caucasian male image; (e) judge the Youth of the Caucasian male image.

It can be seen in the Figure 5.2.2.4 that the poor agreement of the female and male Chinese observers appeared mostly when judging the image with the hue angle of the skin colour equal to 20° and 80°. They also had a poor agreement when judging the *Health* attribute of the image of the OM image with a relatively small hue angle (30° and 40°). The two gender groups gave opposite response to the *Feminine* and *Youth* attributes of the OF images with a hue angle equal to 20° and 80°. The female Chinese observer thought the CF images appeared feminine and youthful when the skin hue angle equal to 20° and appeared masculine and aged when the skin hue angle equal to 80°. The male Chinese observers had the opposite opinion. They also gave opposite responses when judging the *Feminine* and *Youth* attributes of the CM images. The male Chinese observers thought the CM images appeared feminine and youthful when the skin hue was equal to 80°. The female Chinese observers thought the opposite. The male Chinese observers also thought the CM images appeared youthful and aged when the skin hue angle was equal to 70° and 30°. The female Chinese observers held the opposite opinion.

5.2.2.2.3 Correlation between the images with different ethnic groups and different gender

The PCA analysis was used to investigate the relationship between the judgement results of four images (20 sets of data ($4images \times 5attributes = 20$) in total). Each set of data was named with the image and the attribute. For example, the OF (Likeable) indicates the judgement results of the *Likeable* attribute when observing the OF images. Four components were extracted from the data, as listed in Table 5.2.6. Component 1 (marked in red) included all the data sets that were based on the OF images. The data sets, OM (Feminine) and OM (Youth), were also included in Component 1. Component 2 (marked in green) and Component 3 (marked in blue) included all the data sets that were based on the CF and CM images, respectively. Component 4

(marked in orange) included the OM (Sociable), OM (Health) and OM (Likeable) images.

Table 5.2.6 The result of the PCA analysis on investigating the relationship between the judgement results of four images.

	Component			
	1	2	3	4
Rotation Sums of Squared Loadings % of Variance	28.402	23.538	20.355	18.624
OF (Likeable)	.940	-.282	-.044	.064
OF (Feminine)	.934	-.226	-.061	-.019
OF (Youth)	.922	-.254	-.095	-.017
OF (Sociable)	.883	-.299	-.045	.163
OF (Health)	.876	-.275	-.089	.203
OM (Feminine)	.741	-.106	-.017	.384
OM (Youth)	.648	-.054	-.038	.644
CF (Feminine)	-.180	.935	.158	.161
CF (Youth)	-.204	.933	.205	.133
CF (Likeable)	-.318	.920	.204	-.031
CF (Sociable)	-.349	.886	.225	-.103
CF (Health)	-.344	.869	.248	-.096
CM (Likeable)	-.056	.170	.975	-.070
CM (Sociable)	-.047	.141	.929	-.213
CM (Health)	-.048	.163	.893	-.246
CM (Youth)	-.075	.250	.858	.290
CM (Feminine)	-.077	.249	.691	.406
OM (Health)	.015	.022	-.011	.950
OM (Sociable)	.122	.035	-.031	.947
OM (Likeable)	.329	.027	-.024	.931

These four components showed that the correlation between the images with different ethnic groups and genders was weaker than the correlation between attributes, except for the OF images and the OM images. The OM (Feminine) and OM (Youth) images had a stronger correlation to the five data sets of OF images than the other three data sets of OM images. Figure 5.2.2.5 shows the biplots of the four components. Components 1 to 4 are marked in blue, orange, grey and yellow, respectively.

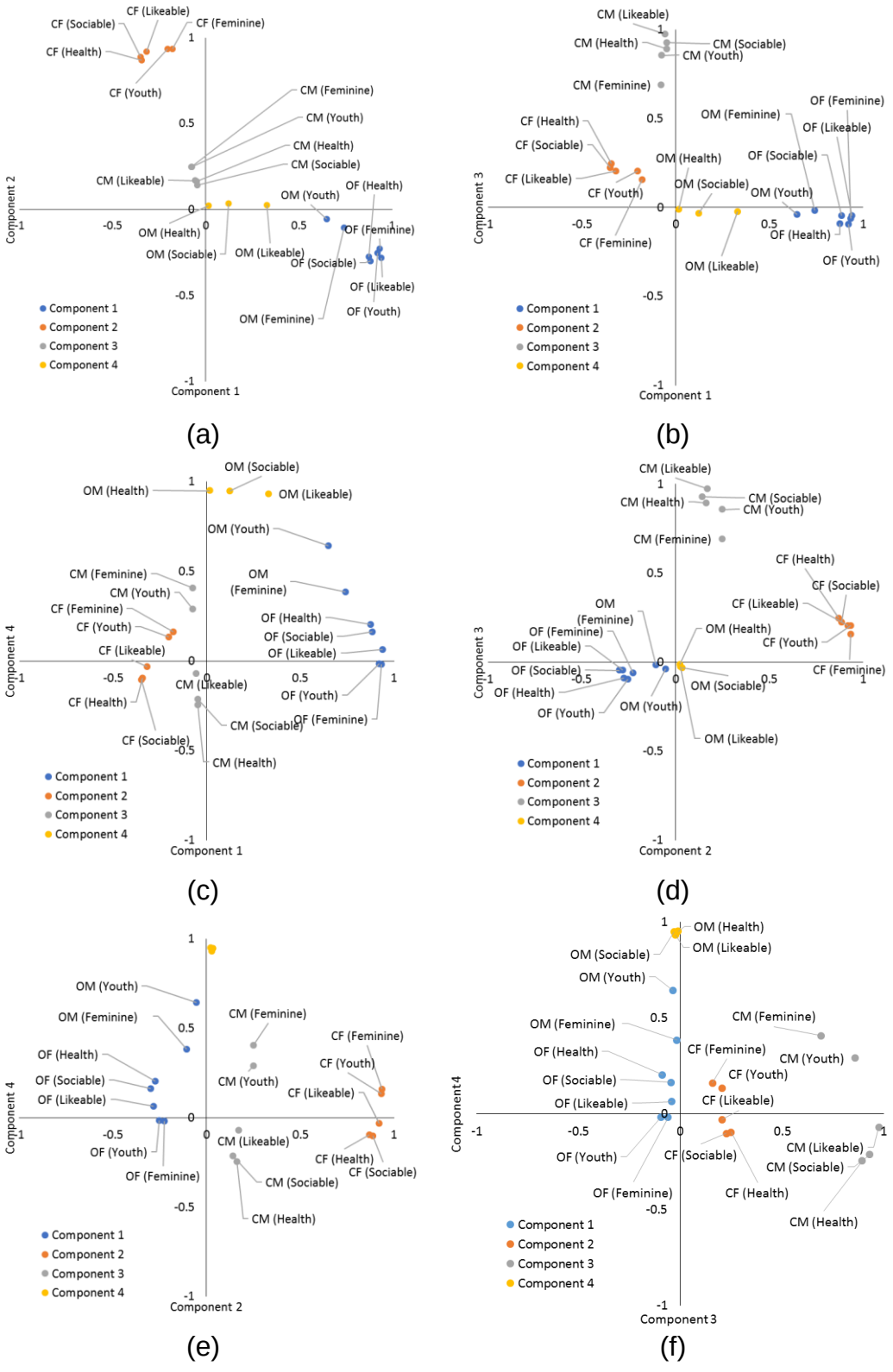


Figure 5.2.2.5 The biplots of the PCA analysis results (Table 5.2.6). (a) Component 1 vs 2; (b) Component 1 vs 3; (c) Component 1 vs 4; (d) Component 2 vs 3; (e) Component 2 vs 4; (f) Component 3 vs 4.

As shown in the figure, four components were distributed scattered. Within the same component, some of the attributes were closely located. This indicates that they had a stronger correlation than the others within the same component. For Component 1, the OF (Feminine) and OF (Youth), OF (Sociable) and OF (Health) were closely located in Figure 5.2.2.5 (a)-(c). Similarly, for Component 2, CF (Feminine) and CF (Youth), CF (Health) and CF (Sociable) were closely located. For Component 3 CM (Sociable) and CM (Health) were closely located. The correlation between the attributes was further investigated in the next section.

The correlation between the images with different ethnic groups and genders were further investigated by using the *r* value. The *r* value of the images with different genders (marked in yellow) and different ethnic groups (marked in green) are listed in Table 5.2.6. The mean *r* values of different ethnic groups and different genders of five attributes are listed in the last column of the same table.

Table 5.2.7 The correlation between two genders and two ethnic groups (*r*).

	Sociable	Likeable	Feminine	Youth	Health	Mean
OM vs OF	0.31	0.36	0.67	0.56	0.27	**0.44**
CF vs CM	0.39	0.38	0.45	0.49	0.43	**0.43**
OM vs CM	-0.18	-0.09	0.12	0.09	-0.18	**-0.05**
OF vs CF	-0.59	-0.57	-0.39	-0.46	-0.57	**-0.52**

The mean *r* value in the last column showed that the judgement results from the same ethnic group but different genders were positively correlated and the mean *r* values were similar. The judgement results from the same gender but different ethnic groups were negatively correlated. The mean *r* values between the OM and CM images were closed to zero (r=-0.05). This implies that the results from these two sets of images had a very weak correlation. The correlation between the OF and CF images had the largest absolute *r* value ($|r|$=0.52). This implies that these two sets of images had relatively strong correlation.

5.2.2.2.4 The correlation between the attributes

The correlation between the judgement results of every two attributes was investigated by using the correlation coefficient, *r*. Table 5.2.8 lists the *r* values between every two attributes for each image, respectively. The mean *r* values of four images at each attribute pairs are listed in the last column of the table. The mean *r* values of all the attribute pairs were listed in the last row of the table. The *r* value, which above or equal to the average r value of all (r=0.82), were marked in yellow.

Table 5.2.8 The correlation between two attributes (*r*).

	OM	OF	CF	CM	Mean
Sociable vs Feminine	0.32	0.84	0.88	0.45	**0.62**
Sociable vs Youth	0.60	0.85	0.90	0.69	**0.76**
Sociable vs Health	0.99	0.99	0.99	0.99	**0.99**
Sociable vs Likeable	0.94	0.96	0.98	0.97	**0.96**
Likeable vs Feminine	0.60	0.95	0.93	0.65	**0.78**
Likeable vs Youth	0.82	0.95	0.95	0.83	**0.89**
Likeable vs Health	0.90	0.96	0.97	0.94	**0.94**
Feminine vs Youth	0.95	0.99	0.99	0.94	**0.97**
Feminine vs Health	0.23	0.83	0.85	0.38	**0.58**
Youth vs Health	0.52	0.84	0.89	0.63	**0.72**
Mean	**0.69**	**0.92**	**0.94**	**0.75**	**0.82**

In Table 5.2.8, the mean *r* values of the four images (last column) show that the correlation between *Sociable* and *Health* attributes, *Sociable* and *Likeable* attributes, *Likable* and *Health* attributes, *Feminine* and *Youth* attributes were stronger (*r* >0.9) than the others. The average *r* values of all the attribute pairs show that the judgement results between attributes of the OF and CF images had a stronger correlation (*r* >0.90) than the others. The *r* values which are greater than 0.82 (the overall mean *r* value) are marked in yellow. This shows that the *r* value of all the attributes of CF and OF images were all higher than the average. Some of the attribute pairs, for example, *Sociable* and *Health*, *Sociable* and *Likeable*, *Likable* and *Youth*, *Likeable* and *Health*, *Feminine* and *Youth*, had *r* values greater 0.82 for all images. Most of these attribute pairs

included *Likeable*. The attribute pair that had an *r* value below the average were the *Likeable* and *Feminine* attributes at OM and CM images. The scatter diagrams for each attribute pairs for each image were plotted and shown in Figure D13-D16 in Appendix D.

5.2.2.2.5 The influence of the skin whiteness and hue angle to the judgement of images

The impact of the hue angle and the skin whiteness on the judgement of these 5 attributes were evaluated via the ANOVA p-value. Through the p-value, the significance of the impact of the skin whiteness and the hue angle on each attribute can be determined. Then a detail investigation was carried out by plotting the judgement results of four images with different skin whiteness and hue angle values in a line chart. The judgement results were unified by transforming them into z-score by using the Thurstone's case V model (Tsukida and Gupta, 2011). The higher z-score value the stronger the response. For example, the *Likeable* attribute had the highest z-score when skin whiteness was equal to 50. This indicates that the facial image appeared the most likeable when the skin whiteness was equal to 50. The influence of both skin whiteness and hue angle on the judgements was also investigated by using a bubble chart (in the next section).

The significance of the impact of skin whiteness and hue angle

The significance of the impact of the skin colour on the judgements of the five attributes was determined by using a one-way ANOVA analysis. The results are listed in Table 5.2.9. A p-value no greater than 0.05 indicates that the scale has a significant impact on the results. As shown in this table, the skin whiteness had a significant impact on the judgement of all five attributes. For the skin hue angle, only the *Sociable* and the *Likeable* attributes were significantly influenced by the hue angle of the skin colour.

Table 5.2.9 The impact of the skin whiteness and hue angle to the 5 attributes (ANOVA p-value ≤ 0.05 indicates as significant).

	Skin whiteness	Skin hue angle
Sociable	0.008*	0.026*
Feminine	0.004*	0.193
Health	2.27E-08*	0.287
Young	9.93E-06*	0.432
Likable	0.012*	0.033*

*indicates a p-value ≤ 0.05

The influence of the skin whiteness of the skin

The z-scores of the images with the same skin whiteness were averaged and plotted in the line chart to investigate the trend of the variation of each attribute, as shown in Figure 5.2.2.6. Four sub-figures are included in this figure, including the line charts for the OF, OM, CF, and CM image sets. In each subfigure, a purple line marks the skin whiteness of the original image. The range of the skin whiteness values of the LLSC database is marked by two black dashed lines. The judgement results of five attributes are plotted in the line charts with different colours.

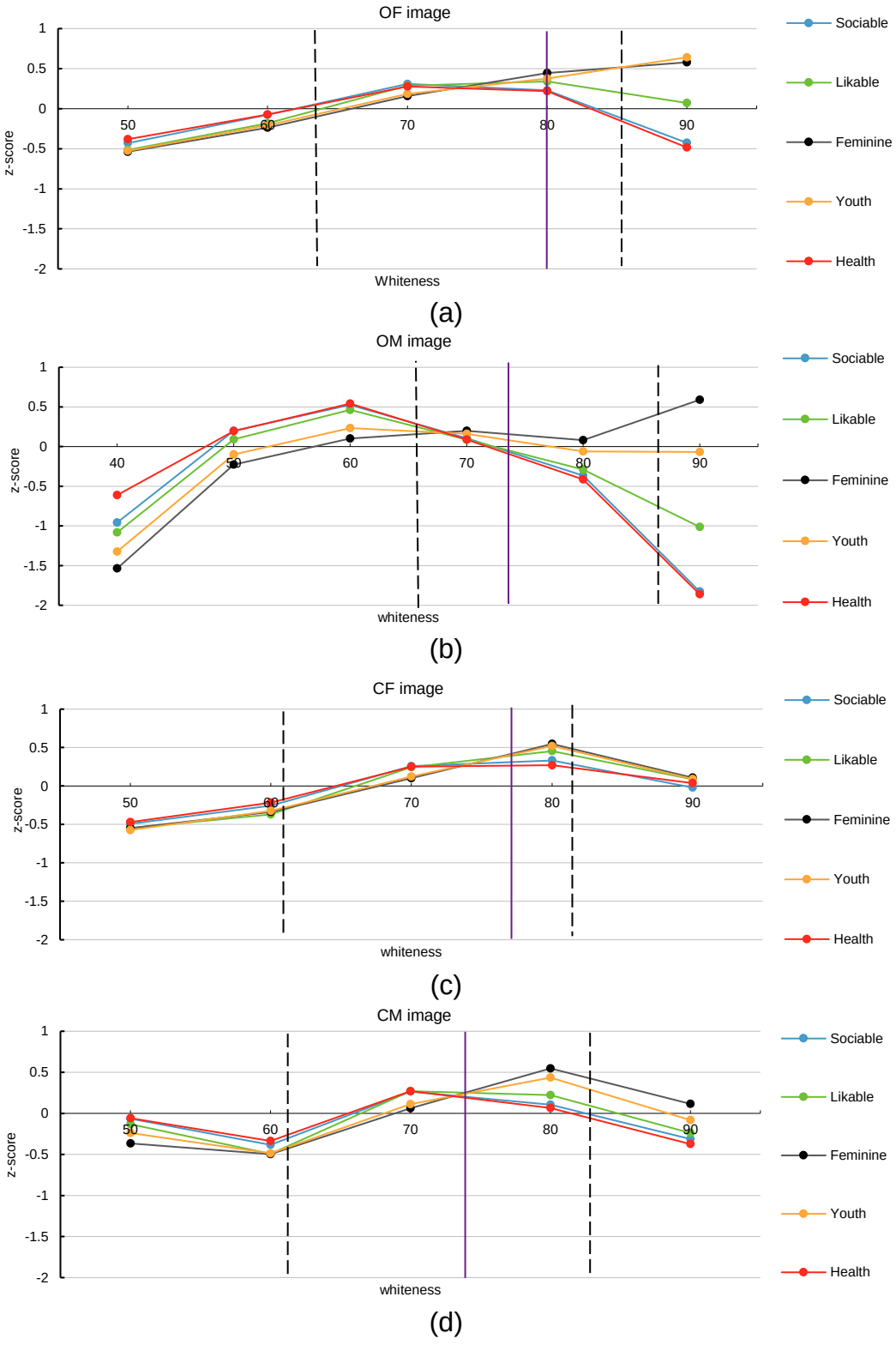


Figure 5.2.2.6 The judgement results of four images with different whiteness. (a) OF images ;(b) OM images ;(c) CF images; (d) CM images.

From these four figures, it can be seen that the *Sociable*, *Likeable* and *Health* attributes had similar variation when judging all four stimuli images. The *Youth* and *Feminine* attribute had a similar distribution to when judging the Oriental female (OF), Caucasian female (CF) and Caucasian male (CM) images.

From Figure 5.2.2.6 (a), it can be seen that the OF images appeared more feminine, youthful, likeable, sociable and healthier when the skin whiteness increased. The increment of the appearance of the likeable, sociable and healthy of the OF images meet a plateau when the skin whiteness is 70. Then, they drop as skin whiteness approaches 90. The original OF image has a skin whiteness of 74.99. For healthier and more sociable appearance within the LLSC skin colour gamut, the skin whiteness needs to be reduced to 70. For more likeable, feminine and youthful appearance within the LLSC skin colour gamut, the skin whiteness needs to increase to 80.

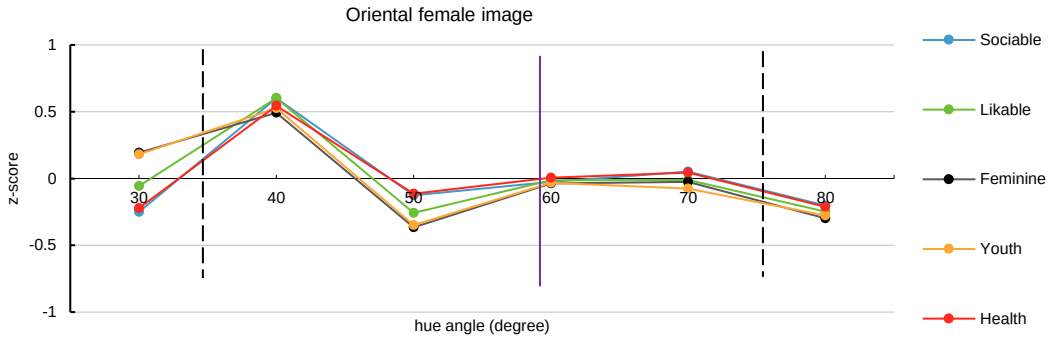
For the OM image, as shown in Figure 5.2.2.6 (b), the trend of the judgements with increasing skin whiteness is close to the shape of a bell, except for the *Feminine* attribute. The OM images appeared to be the most sociable, likeable, youthful and healthiest when the skin whiteness is equal to 60. The *Feminine* attribute of the OM images increased when the skin whiteness increased. The skin whiteness of the original image was 68.83. For more sociable, likeable, healthier and younger appearance within the LLSC skin colour gamut, the skin whiteness needs to at least 60. For a more masculine appearance within the LLSC skin colour gamut, the skin whiteness needs to reduce to 60.

For the CF image, as shown in Figure 5.2.2.6 (c), the trend of the *Sociable, Likeable*, *Feminine*, *Youth* and *Health* attributes is similar to that of the OF images. Except the Feminine and Youth attributes drop when the skin whiteness is equal to 90. The skin whiteness of the original image was 79.02. To achieve healthier appearance within the LLSC skin colour gamut, the skin whiteness needs to be reduced to a value of 70. The skin whiteness needs to increase to 80 to achieve the most sociable, likeable, feminine and youngest appearance within the LLSC skin colour gamut.

For the CM image, as shown in Figure 5.2.2.6 (d), the trend of the five attributes with increasing skin whiteness is similar to that of the CF image, except the CM image appeared more autistic, masculine, aged, dislikeable and unhealthy at a skin whiteness of 60 than at a skin whiteness of 50. The skin whiteness of the original image was 75.64. To achieve a younger appearance within the LLSC skin colour gamut, the skin whiteness of the original image needs to increase to a value of 80. For more sociable, likeable and healthier appearance, the skin whiteness needed to reduce to a value of 70.

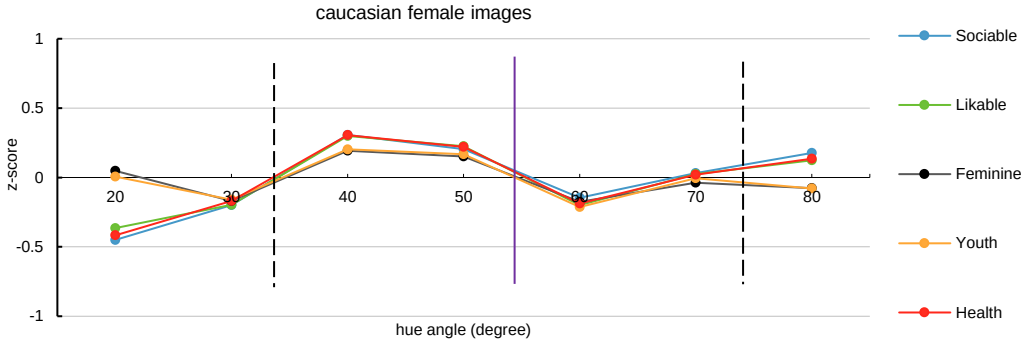
The influence of the hue angle of the skin

The z-scores of the images with the same hue angle were averaged and plotted in a line chart, as shown Figure 5.2.2.7. The range of the hue angle of the LLSC skin colour gamut is marked by two black dashed lines. The hue angle of the original image is marked by a purple line.

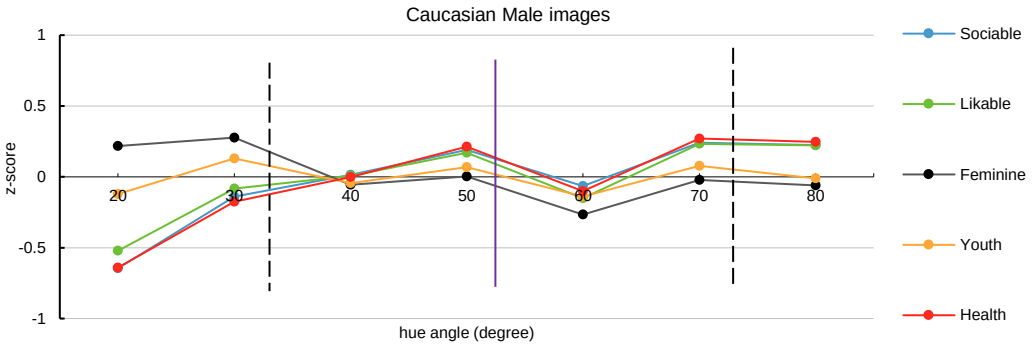


(a)

(b)



(c)



(d)

Figure 5.2.2.7 The judgement results of four images with different hue angles. (a) OF images; (b) judgement of observing OM images ;(c) judgement of observing CF images; (d) judgement of observing CM images.

For the OF images, the trend lines for the 5 attributes show a similar trend and agreed on response at different values of hue angle values, as shown in Figure 5.2.2.7 (a). But, for the hue angle of the skin colour equal to 30°, the image appeared to be autistic, dislikeable, unhealthy feminine and youthful. The hue angle of the skin colour in the original image was 59.80°. The hue angle of the original skin colour needed to be reduced to 40° to achieve more sociable, likeable, feminine younger and healthier appearance within the LLSC skin colour gamut.

For the OM image, the trends of the 5 attributes show a similar trend with increase in hue angle, as illustrated in Figure 5.2.2.7 (b). The OM images appeared less sociable, likeable, youthful, feminine and healthy when the hue angle of the skin colour was increased. The hue angle of the original image was 58.84˚. The hue angle of the original image needs to be reduced to 40° to achieve more likeable, sociable and healthier appearance within the LLSC skin colour gamut. To achieve younger appearance, the hue angle of the original image needs to be reduced to 50°. For a more masculine appearance, the hue angle of the original skin colour needs to increase to 60°.

For the CF image, as shown in Figure 5.2.2.7 (c), the trend of *Health*, *Sociable* and *Likeable* attribute were similar. The trend of the *Youth* and *Feminine* attributes were also similar. The trends of these five attributes were bell-shaped. The hue angle of the skin colour of the original image was 51.47˚. The hue angle of the skin colour needs to be reduced to 40° to achieve a more sociable, likeable feminine healthier and younger appearance within the LLSC skin colour gamut.

For the CM image, the trends of the five attributes were similar. They all had oscillating results, as shown in Figure 5.2.2.7 (d). This could be because of the Chinese observers had only a limited knowledge of the typical skin colour of a Caucasian male. Caucasian females frequently appear in the TV advertisements but the appearance of Caucasian males is limited. This could cause the Chinese observers to give poor judgements of the CM images with different hue angles. The hue angle of the skin colour of the original image

was 51.68°. The hue angle of the original image needs to be reduced to 50° to achieve a more sociable, likeable and healthier appearance within the LLSC skin colour gamut. For younger appearance, the hue angle needs to be reduced to 30°. For more masculine appearance, the hue angle needs to be increased to 60°.

Influence from whiteness and hue angle of the skin colour

The influence of both skin whiteness and hue angle on the judgements of the five attributes was investigated by illustrating the data on a bubble chart. The variation in the response to each attribute when observing the 26 OF images, the 26 CF images, the 26 OM images and the 26 CM images are shown in Figure 5.2.2.9, Figure 5.2.2.9, Figure 5.2.2.10 and Figure 5.2.2.11, respectively. The bubbles represent the z-score values of each image for each attribute. The blue bubble in the figure represents a positive z-score and the white bubble represented the negative z-score. The size of the bubble is an indicator of the absolute value of the z-score. Thus, the larger size of the bubble means the absolute value of the z-score was larger. The red dashed lines in the figure indicate the range of the skin colour in terms of the skin whiteness and hue angle based on the LLSC database. The rectangular area within the red lines, which is marked in green, is the range of the skin colour within the LLSC database. The z-scores of the original images are marked in orange.

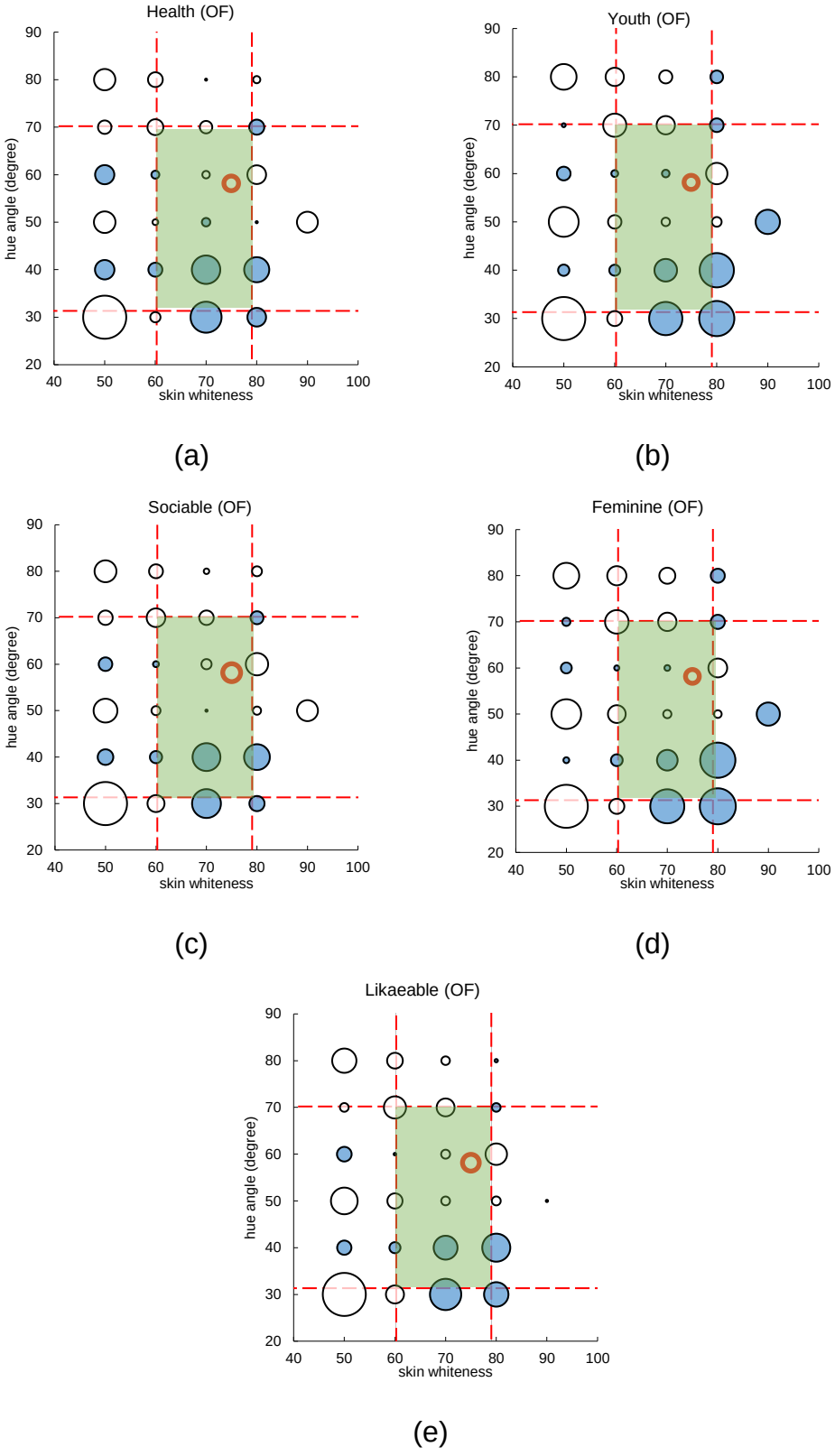


Figure 5.2.2.8 The bubble chart of judging 26 OF images (original image marked in orange) at: (a) Health; (b) Youth; (c) Sociable; (d) Feminine; (e) Likeable.

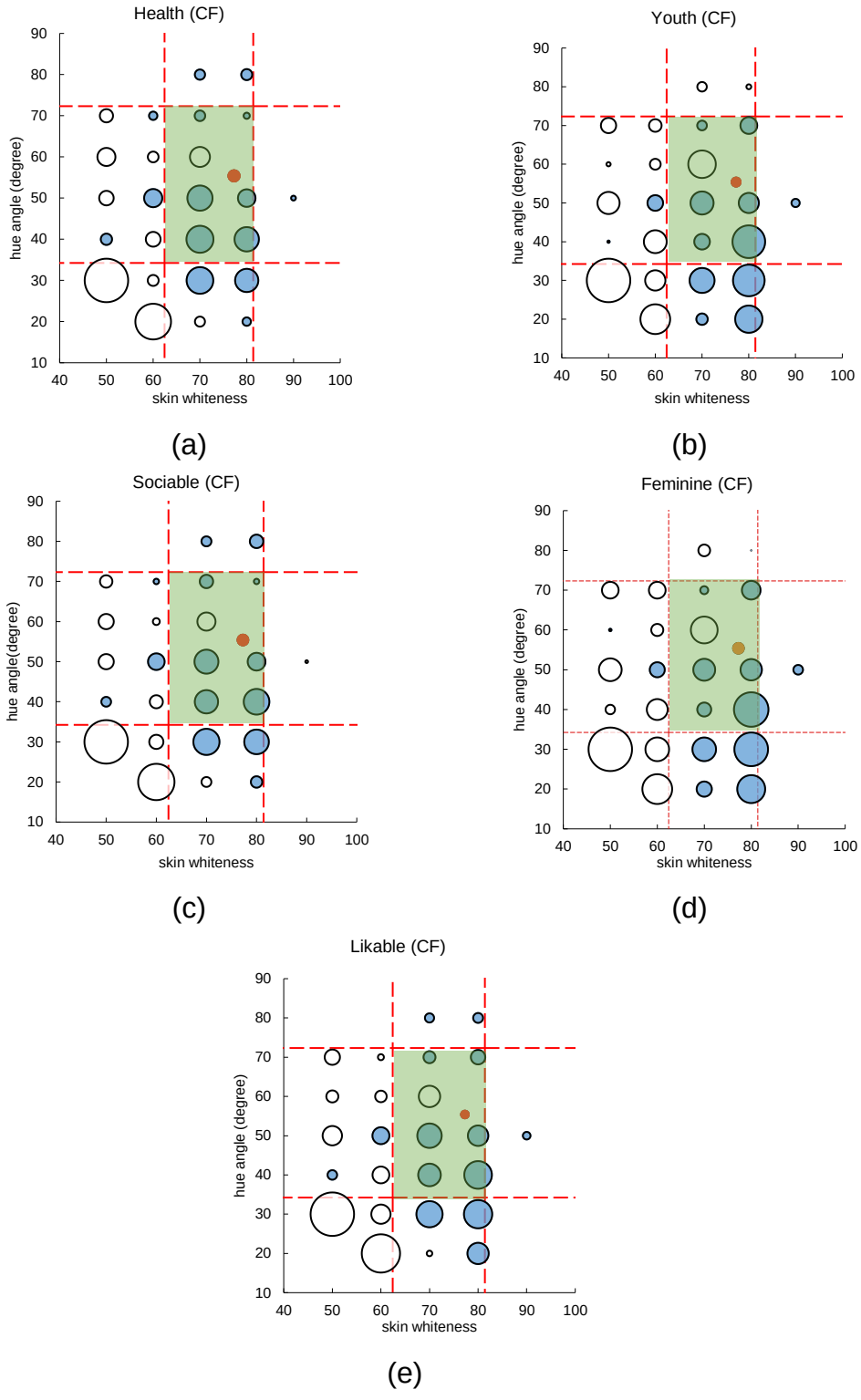


Figure 5.2.2.9 The bubble chart of judging 26 CF images (original image marked in orange) at: (a) Health; (b) Youth; (c) Sociable; (d) Feminine; (e) Likeable.

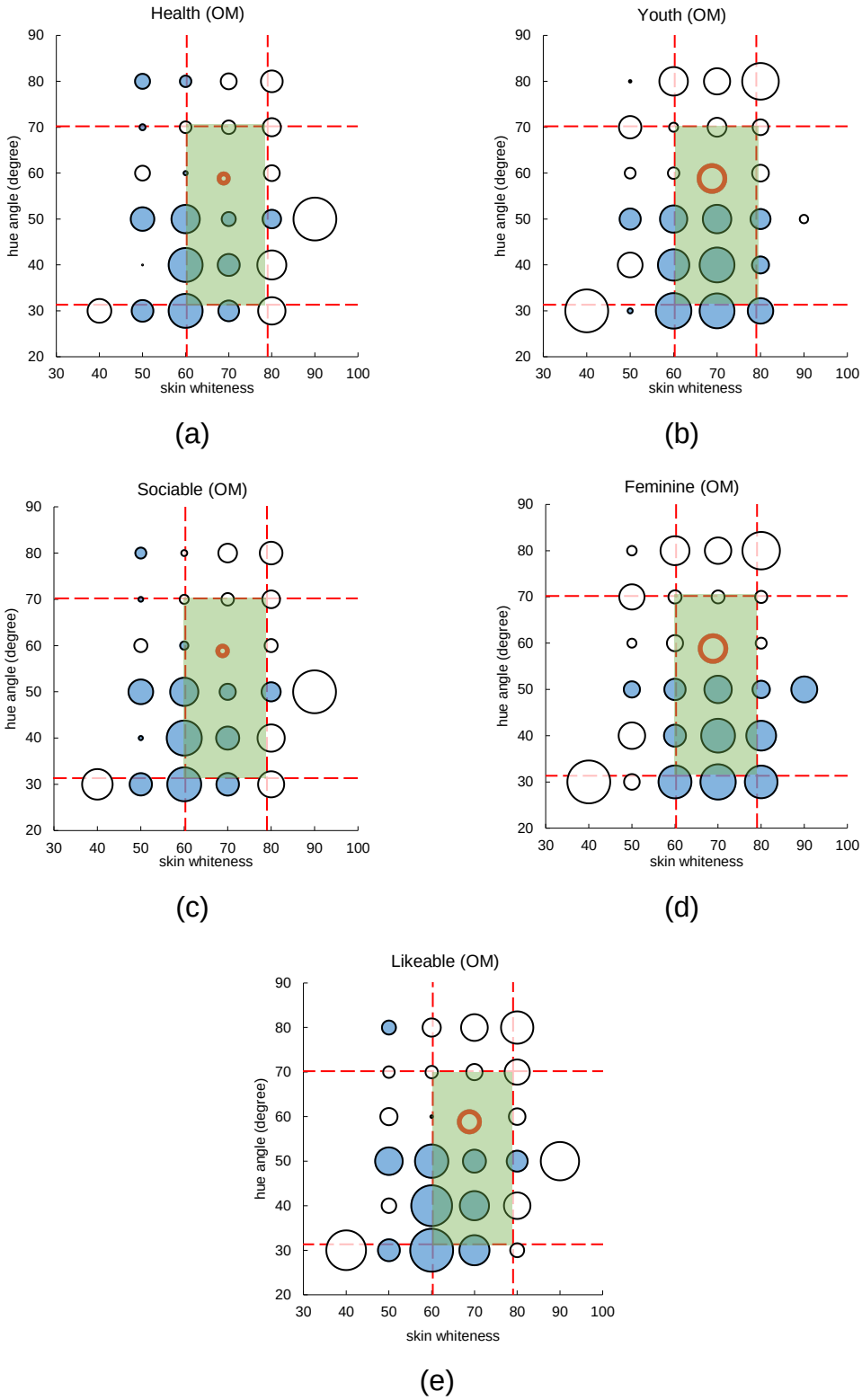


Figure 5.2.2.10 The bubble chart of judging 26 OM images (original image marked in orange) at: (a) Health; (b) Youth; (c) Sociable; (d) Feminine; (e) Likeable.

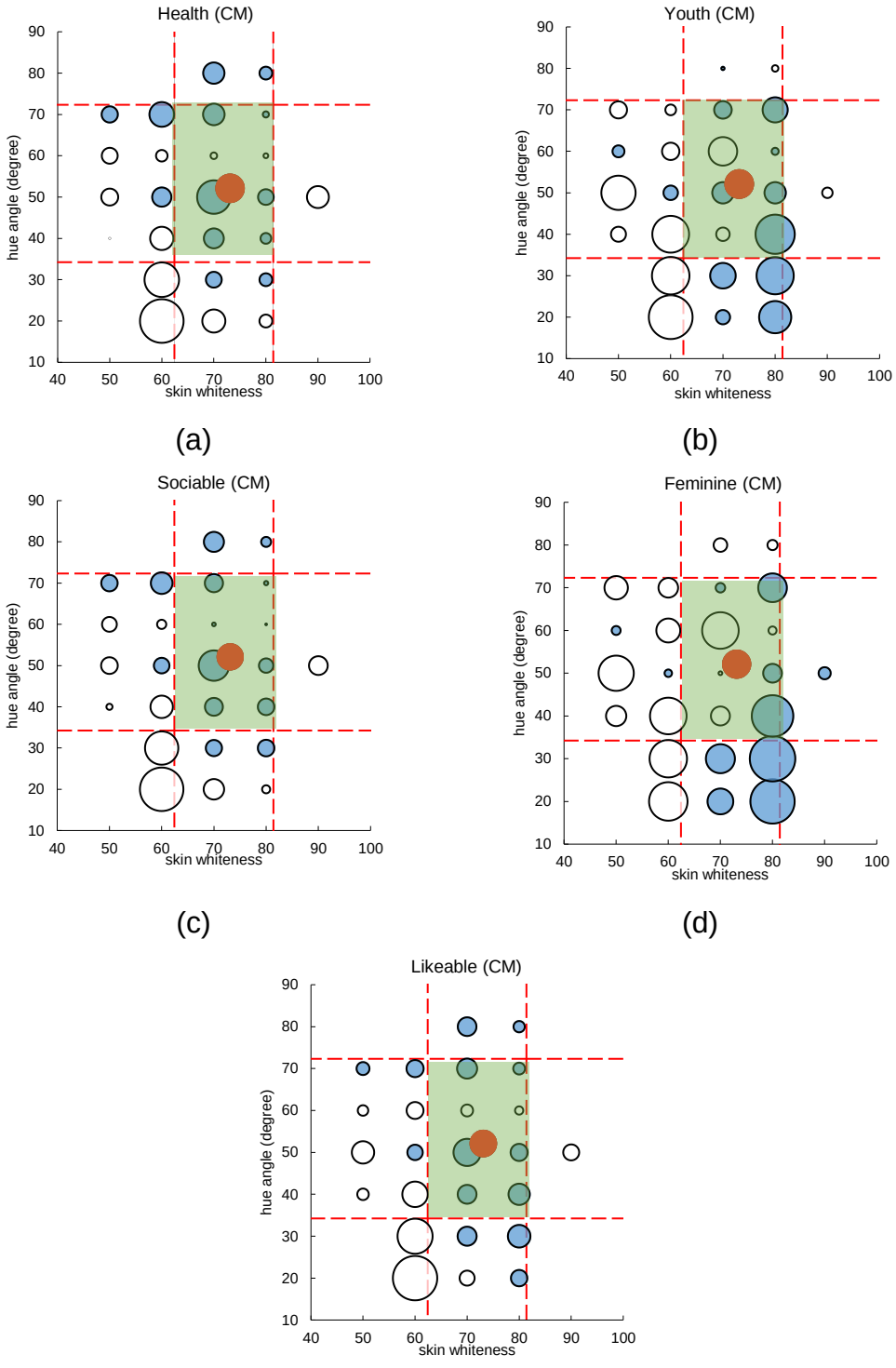


Figure 5.2.2.11 The bubble chart of judging 26 CM images (original image marked in orange) at: (a) Health; (b) Youth; (c) Sociable; (d) Feminine; (e) Likeable.

From Figure 5.2.2.9 to Figure 5.2.2.11, some consistent trends can be found as follows

- For the OF and the CF images, the strong positive responses to the five attributes appeared at the high skin whiteness (70 to 90) and low hue angle area (OF:30˚ to 40˚; CF: 30˚ to 50˚). This indicates that the whiter and redder skin colour were appeared more feminine, likeable, sociable, youthful and healthier. This confirms results found by Yano, Yoshikawa and Suzuki (1997) (see sections 2.2.3.1).

- For the OM images, the strong positive responses to the Health, Likeable, Sociable and Youth attributes mostly appeared at a medium skin whiteness (60 to 70) and low hue angle (30˚ to 50˚) area. This shows that the reddish and tan OM skin colour appeared more likeable, sociable, youthful and healthier. For the masculine appearance of the OM images, the skin colour needs to be yellowish.

- For the CM images, the positive responses of the *Health*, *Sociable* and *Likeable* attributes were mostly distributed in two sections: 1. the skin colour with skin whiteness from 50 to 70 and hue angle above 70°; 2. the skin colour with skin whiteness from 70 to 80 and hue angle from 50° to 40°. This suggests that the dark yellowish and white reddish skin colours of the CM images both appeared more sociable, likeable and healthier. The response for the *Youth* and *Feminine* attributes was similar to that for the CF images.

Comparing the present study to others on preferred skin colour

The present study was compared with the previous work of three researchers which studied the skin colour preference (see Section 2.2.3.1.). Table 5.2.10 summarises this previous work and the findings of the present study. The weighted mean of the results shown in Figure 5.2.2.8 (e) (OF-Likable) was used for comparison. The weighted mean of four images, which had the z-

score values as the top four, was calculated and is listed in the Table 5.2.10. Equation 5.5 is used to complete the calculation.

$$CIELAB_{w.mean} = \frac{\sum_{i=1}^{n} L^* a^* b^*{}_i z_i}{\sum_{i=1}^{n} z_i} \qquad \textbf{Equation 5.5}$$

where $L^* a^* b^*{}_i$ is a matrix (3×1) contained the CIELAB values of the skin colour of the image i. z_i is the z-score of the image i. and n is the number of the images included in the calculation.

Table 5.2.10 Skin colour preference for previous and the present studies (F and M were shorted for female and male).

	Stimuli Image	**Observer**	C^*_{ab}	h_{ab} (degree)
Zeng and Luo (2011)	Chinese-F (D50)	Chinese-F&M	30.0	51.1
Yano and Hashimoto (1997)*	Japanese-F (D65)	Japanese-F	23.4	39.9
Fan *et al*. (2011)	Japanese-F (D50)	Japanese-F&M	22.2	42.1
Fan *et al*. (2011)	Japanese-F (D50)	Chinese-F&M	21.8	35.8
present study	Chinese-F (D65)	Chinese-FM	21.8	37.8

Yano reported their finding in the u'v' chromaticity space (0.2425, 0.4895). These data were transformed into CIEXYZ, then CIELAB by assuming L=60. This transformation is explained in Appendix B.

As shown in the table, the present study has good agreement with the work of Yano (1997) and Fan (2011), where the reddish skin colour was more preferred. This finding is also agreed with the finding of other researchers, *e.g.* Yoshikawa *et al*. (2010). Yoshikawa *et al*. also found that the whiter and reddish skin colour was more preferred by Japanese observer. The findings of Zeng and Luo not well agree with the present study. The preferred skin colour that found by Zeng and Luo was yellower in hue, comparing with the present study. But Zeng and Luo reported that the preferred skin colour was redder than the actual skin colour. Note that Zeng and Luo used images with different a^* and b^* values and no variation in L^* was not examined.

5.2.2.3 Modelling the results from Experiment 2.2

This section is aimed to develop the models, which can predict five perceived impression via skin colour (skin whiteness and hue angle), by using the experimental results of Experiment 2.2. Here, the models were developed in the form of polynomial functions. Twenty models were built to simulate the judgement results for the two ethnic groups and the two genders for five attributes ($2\ \text{genders} \times 2\ \text{ethnic groups} \times 5\ \text{attributes} = 20$). The *MATLAB Curve Fit Toolbox* was used to fit the data and derive the models. Different degrees of the independent variable were tested, and the one with the best fit (judged by the r^2 value) was selected. Equation 5.6 is the final formula of the model.

$$Model = p00 + p10h + p01w + p20h^2 + p11hw + p02w^2 + p30h^3 + p21h^2w + p12hw^2 + p03w^3$$

Equation 5.6

where $h = \cos(h_{ab})$, and w is the skin whiteness value. w can be calculated by using Equation 4.1. The p00, p10, p01, p11, p02, p30, p21, p12 and p03 were the constant coefficients. Different models have different coefficient values.

Table 5.2.11 lists the coefficients of each model. The accuracy of the model was also tested by using the experimental results of Experiment 2.2. The modelling results were plotted against the visual results, as shown in Figures D16 to D17 in Appendix D. The *r* between the estimated results and experimental results were calculated to determine the accuracy of the model as listed in Table 5.2.11.

Table 5.2.11 The constant coefficients and the test results of the models.

		Coefficients										
		p00	**p10**	**p01**	**p20**	**p02**	**p30**	**p03**	**p11**	**p21**	**p12**	**Test Result(r)**
OF	**Sociable**	1.1783E+01	-6.8762E+00	-5.5687E-01	-4.5698E+00	7.9714E-03	-1.4052E+01	-3.2094E-05	3.1203E-01	3.8549E-01	-4.9509E-03	0.79
	Likeable	7.4710E+00	1.8195E+00	-3.7559E-01	-1.6565E+01	4.9655E-03	-1.0896E+01	-1.4092E-05	2.0290E-01	5.0566E-01	-4.9437E-03	0.80
	Feminine	2.2123E+00	1.0608E+01	-1.4274E-01	-3.3920E+01	9.1042E-04	-1.4265E+00	9.5134E-06	1.5696E-01	5.5919E-01	-4.8952E-03	0.87
	Youth	-4.6933E+00	1.2913E+01	1.5713E-01	-3.2652E+01	-3.2687E-03	-3.4637E+00	2.8885E-05	7.2564E-02	5.8925E-01	-4.5123E-03	0.86
	Health	1.2761E+01	-5.7501E+00	-6.0139E-01	-9.2005E+00	8.3965E-03	-1.2140E+01	-3.2454E-05	3.5616E-01	4.0844E-01	-5.5006E-03	0.79
OM	**Sociable**	1.9007E+01	-5.6877E+01	-5.7863E-01	2.8115E+01	5.0004E-03	-9.3708E+00	-9.1620E-06	1.4296E+00	-2.0661E-01	-9.6110E-03	0.94
	Likeable	1.1075E+01	-4.9531E+01	-2.2006E-01	2.1440E+01	-5.0002E-04	-6.6179E+00	1.8404E-05	1.2652E+00	-1.5973E-01	-8.4740E-03	0.92
	Feminine	8.7500E+00	-3.0143E+01	-1.7560E-01	2.8306E-01	-1.2546E-03	-6.5542E-01	2.4393E-05	9.2645E-01	1.9921E-02	-6.7145E-03	0.93
	Youth	1.1151E+01	-3.6741E+01	-2.6775E-01	7.8943E+00	1.7917E-04	-3.6373E+00	1.6871E-05	1.0459E+00	-2.6902E-02	-7.5090E-03	0.92
	Health	2.1292E+01	-5.9489E+01	-6.6279E-01	3.4161E+01	6.3717E-03	-1.1191E+01	-1.7123E-05	1.4024E+00	-2.4472E-01	-9.1161E-03	0.92
CF	**Sociable**	9.8438E+00	-1.5932E+01	-3.8725E-01	1.5607E+01	5.5702E-03	-2.3145E+01	-2.3108E-05	2.4209E-01	3.0166E-01	-3.5731E-03	0.89
	Likeable	1.2722E+01	-4.8930E+00	-5.9822E-01	7.5460E+00	9.3723E-03	-2.0875E+01	-4.3633E-05	3.2575E-02	3.6369E-01	-2.4528E-03	0.89
	Feminine	1.9721E+01	-8.2494E-01	-9.1453E-01	-7.5267E+00	1.3157E-02	-9.5998E+00	-5.7599E-05	1.3770E-01	3.2135E-01	-2.9873E-03	0.89
	Youth	1.7458E+01	-2.3607E+00	-8.0016E-01	-4.7253E+00	1.1558E-02	-1.1319E+01	-5.0796E-05	1.3006E-01	3.2307E-01	-2.8728E-03	0.87
	Health	9.1850E+00	-2.1981E+01	-3.0822E-01	2.0476E+01	3.9454E-03	-2.2991E+01	-1.4166E-05	3.4591E-01	2.3071E-01	-3.8224E-03	0.87
CM	**Sociable**	2.1590E+01	-5.7059E+00	-9.8348E-01	1.6371E+01	1.6148E-02	-1.4563E+01	-8.8055E-05	-1.4250E-01	7.8892E-02	1.1489E-03	0.89
	Likeable	2.5185E+01	-1.5447E+00	-1.1708E+00	1.0314E+01	1.8940E-02	-1.3596E+01	-1.0040E-04	-1.8530E-01	1.5135E-01	9.6889E-04	0.89
	Feminine	2.0753E+01	2.1027E+01	-1.1253E+00	-1.6948E+01	1.8474E-02	6.8073E-01	-9.5049E-05	-4.2259E-01	2.3652E-01	1.7793E-03	0.88
	Youth	2.6057E+01	1.3781E+01	-1.3184E+00	-8.0652E+00	2.1292E-02	-3.4440E+00	-1.1034E-04	-3.3925E-01	1.8498E-01	1.6565E-03	0.84
	Health	1.9458E+01	-5.4910E+00	-8.9798E-01	1.6544E+01	1.5004E-02	-1.3971E+01	-8.3014E-05	-1.4175E-01	6.2269E-02	1.1806E-03	0.87
											Average	0.88

The *r* values in Table 5.2.11 showed that the visual results and the modelling results were above 06 which showed a strong correlation. This indicates that the models have good accuracy at estimated these test data.

5.2.2.4 Summary of Experiment 2.2

Experiment 2.2 used the *Paired Comparison* method to examine the impact of skin colour on the five attributes which were selected via Experiment 2.1. The influence of the gender and ethnic group of the stimuli images, and the influence of the gender of the Chinese observers on the judgement were also analysed in this section.

The female and male Chinese observers have good agreement when judging the five attributes with images that varied in skin whiteness. Good agreement was also found when judging the five attributes at the OF images and the OM images with different skin colours as defined by the hue angle. The agreement of judging the Feminine and Youth attributes of the CF and the CM images with different skin colour hue angle were poorer than the others.

The significance of the impact of the skin whiteness and hue angle was investigated by using an ANOVA analysis. These results showed that the skin whiteness has a significant impact on the judgement of all five attributes. The hue angle, however, only has a significant impact on the judgement of the *Sociable* and *Likeable* attributes.

For the influence of skin whiteness, the OF and CF images appeared more sociable, likeable, feminine, youthful and healthier with high skin whiteness; the OM images appeared more sociable, likeable, masculine, youthful and healthier with medium skin whiteness; the CM images appeared more youthful and less masculine with high skin whiteness and appeared more sociable, likeable and healthier with skin whiteness equal to 70.

For the influence of the hue angle, the OF and CM images appeared to be the most sociable, likeable, feminine, youthful and healthiest with the hue angle

of the skin colour equal to 40º; the OM images appeared to be the most sociable and healthiest with the hue angle of the skin colour equal to 40º, and appeared to be the most likeable, feminine and youthful with the hue angle of the skin colour equal to 50º; the CM images appeared to be the most sociable, likeable, and healthiest with the hue angle of the skin colour equal to 50º and more feminine, youthful with the hue angle of the skin colour equal to 70º.

The distribution of the judgement results in the skin whiteness and hue angle plane reflected the same trends that were found for the skin whiteness and hue angle. The judgement results for the OF image for the likeable attribute had good agreement with the finding of Japanese researchers for skin colour preference.

A polynomial model was used to simulate the judgement results. It shows acceptable accuracy when tested test by using a training sample set.

5.3 Summary of Chapter 5

This chapter presents two experiments that investigated the influence of the skin colour on the impression of facial images. Skin colours within and beyond the LLSC skin colour range were selected to ensure the maximum range of skin colour was investigated. A newly developed skin whiteness index was used in the present study to reduce the complexity of the skin colour selection and analysis.

Two psychophysical experiments that were carried out in the present study were named Experiment 2.1 and Experiment 2.2. Experiment 2.1 used 52 images to investigate the relationship between 23 attributes when judging the images with different skin colours. The correlation coefficient and principal component analysis (PCA) was used to analyse the data. Through PCA analysis, the number of the attributes was reduced from 23 to 5. These five attributes were *Sociable*, *Likeable*, *Feminine*, *Youth* and *Health*. Based on

Experiment 2.1, Experiment 2.2 used 104 images, including those of two ethnic groups with both genders, to investigate the influence of skin colour, the gender of the Chinese observer and their cultural background, on the judgement of five attributes. It was shown from the ANOVA analysis that skin whiteness has a significant impact on the judgement of all five attributes. The hue angle of the skin colour has a significant impact on judging the *Sociable* and *Likeable* attributes. Further analysis of the influence of the skin whiteness and skin hue angle on the judgement of the five attributes was carried out using a plot of the mean z-scores of the measurement results in the line charts. Through the trend of the lines in the line chart, the impact of the skin whiteness and hue angle on the judgements of the images with different gender and ethnic group at different attributes could be determined.

For the impact of the skin whiteness:

- The trend of the judgement results of the Feminine and Youth attributes were similar to the judgements of the OF and CF image sets.

For male images, the ethnic group of the image greatly influenced the judgement results. For the OM and the CM images, there were different trends for all five attributes. For female images, the ethnic group of the image had a limited influence on the judgement results. The OF and CF images had a similar trend for all five attributes when the skin whiteness range was within the LLSC skin colour range.

For the impact of the skin hue angle:

- The variation of the judgements from observing the images with different skin hue angles was appeared different.
- No clear trend was found in the variation of the judgements when observing the CM images with different skin hue angles.

The ethnic group and gender of the subject in the stimuli image had an impact on the judgements. But both the OF and the CF images with a hue angle of 40° appeared to be more sociable, likeable, feminine youthful and healthier than the other attributes.

The gender of the Chinese observer was found to have an impact on the judgements. Two gender groups had a poor agreement when at judging the OF images with different skin whiteness and when judging the CF and CM images with different skin hue angle. When judging different attributes, the two gender groups had a poor agreement when judging the *Feminine* and *Youth* attributes of the images with different skin hue angle.

Models were built and tested using the experimental results from Experiment 2.2. The accuracy each model was tested by use of the training data and found to be good. The attributes were also tested by using each model on other attributes. As to be expected the accuracy was not as good as using the correct model. Only the model of the Likable attribute showed good accuracy on estimated values with the models from other attributes.

Chapter 6 Conclusion

6.1 The study findings

As described in Section 1.2, the overall goal of the present study was to investigate the influence of skin colour on the judgement of the perceived facial impression attributes. To achieve this goal, two objectives are set. The first objective was to establish a comprehensive skin colour database. This skin colour database should include the skin colour of a large number of subjects from different ethnicities, and different instruments should be used to measure different body and facial locations for each subject. The specific tasks included to:

- Determine the measurement protocols of the two instruments (TSR and SP) by investigating of the short-term repeatability of each instrument for measuring human skin colours in vivo;
- Understand the distribution of the skin colours of different locations of a subject, of different ethnic groups, of different genders and of different instruments;
- Formulate the relationship between skin colours and colorimetric data.

The second objective was to study the impact of the skin colour on the perceived facial impression attributes that frequently and popularly used. The tasks included to:

- Accumulate perceived facial impression attributes,
- Conduct the psychophysical experiments to find the basic dimensions to describe facial impression,
- Develop the models to simulate the relationships between the colorimetric values and the perceived facial impression attributes for each ethnic and gender group.

All the targets set at the beginning of the project were successfully accomplished. This chapter summarises the findings.

6.1.1 The findings from Experiment 1 LLSC database

A skin colour database was established and was named the Leeds Liverpool skin colour database (LLSC). It included skin colour of ten locations from 188 subjects, including four ethnic groups (Caucasian, Oriental, South Asia and African) with both genders. Three measurement methods were used, including a Photo Research PR650 tele-spectroradiometer, a Konica-Minolta CM700d spectrophotometer (SP), visual assessments (VA) based on a set of skin colour chart. Each method was used to measure the fixed locations for each subject. The results were reported in terms of spectral reflectance (400 nm to 700 nm, 1 nm interval). A digital camera was used to capture facial images.

6.1.1 Determine of the instruments

The standard measurement procedure of the two instruments (PR650 and CM700d) were determined through a series of the short-term repeatability test (Section 3.1.2.4). The results are summarised below:

Both instruments have better short-term repeatability at measuring colour patches (MCDM value below 0.1 ΔE^*_{ab}) than that of measuring real human skin colour in vivo (MCDM value about 0.5 ΔE^*_{ab}).

The CM700d had better short-term repeatability (0.07 ΔE^*_{ab} and about 0.40 ΔE^*_{ab}) than that of the PR650 (0.09 ΔE^*_{ab} and about 0.50 ΔE^*_{ab})at measuring colour patches and at measuring human skin colour in vivo, respectively.

The CM700d had better short-term repeatability when using continuous repeatability measurement (CT) method which kept the instrument contact the target area during the repeat measurement. The short-term repeatabilities of using different masks were similar. The MAV-LP (Φ 8 mm with larger contact area) was slightly better than the others (0.35 ΔE^*_{ab}).

The PR650 had same short-term repeatability at different measurement distance. But the colour difference between the measurement results of the PR650 with measurement distance of 57.5 cm (P1) and that of the CM700d (MAV-LP) was the smallest (2.57 ΔE^*_{ab}), as the size of the measurement area was similar to the MAV(measuring field size PR650 at P1: Φ 10 mm).

So the measurement protocols of two instrument were: the measurement distance of the PR650 was 57.5 cm (P1); the CM700d used MAV-LP mask and measured the target location used CT method.

These measurement protocols were used to accumulate skin colour data. The measurement results of two instruments were also analysed to investigate the short-term repeatability of the same location but of the different subjects in each ethnic groups. The results are summarised below.

It was found , for the same five locations (forehead, cheek, cheekbone, neck and the back of the hand), the typical short-term repeatability were 0.89, 0.71, 0.76 and 1.00 ΔE^*_{ab} for the Caucasian, Oriental, South Asia and African groups of the PR650 results, respectively. Similarly, the results were 0.43, 0.36, 0.43 and 0.20 ΔE^*_{ab} for the Caucasian, Oriental, South Asia and African of the CM700d results, respectively. The results clearly showed that CM700d has better short-term repeatability at measuring human skin colour in vivo.

Different measurement methods and different ethnicities

For the mean spectral reflectance of the each of the four ethnic groups of two instrument measurement results, in general, the shape of the spectral reflectance from the two instruments (PR650 and CM700d) had good agreement for each ethnic group. Also, the difference between different ethnic groups is mainly in the magnitude but not in shape, except the African group.

All the skin colour of the LLSC database were plotted in CIELAB a*b* and $L^*C^*_{ab}$ planes to understand the colour distribution of measurement methods. The mean CIELAB of the skin colour of each ethnic of each measurement method were plotted in CIELAB a*b* and $L^*C^*_{ab}$ planes to understand the

difference between different measurement methods and between different ethnicities. The results can be summarised as below:

- The measurement results of different measurement methods had similar shape of distribution in a^*b^* and $L^*C^*_{ab}$ plane. The distribution of the measurement results in the $L^*C^*_{ab}$ plane was similar to the shape of a banana.
- The ranges of the hue angle of different measurement methods were similar (Oriental:about 40˚, Caucasian: about 40˚, South Asian: about 20˚, African: about 15˚). Note that the number of participants in South Asian and African groups are small. The hue angle range of measurement results of the CM700d is slightly larger than the other methods. because the results of CM700d included the skin colour of 10 measurement locations, the other method only included the skin colour of 4 locations.
- The PR650 and CM700d gave similar measurement results for the four ethnicities studied. VA method gave the measurement results with higher lightness value and hue angle value than those from the two instruments.
- All the measurement methods gave the results that the Caucasian had higher lightness, followed with Oriental, South Asian and African in the order.

Difference between locations

The difference between different locations was investigated by plotting the mean CIELAB of each location of each ethnicity in CIELAB a^*b^* and $L^*C^*_{ab}$ planes. For the Caucasian, Oriental and South Asian groups, the locations exposure less to the sun, were lighter and less colourful. The body locations were yellower than the facial locations. For the African group, the differences between different locations were mainly in chroma values and lightness values.

Difference between genders

The difference between genders was investigated via theΔL^*, ΔC^*_{ab} and Δh_{ab} (mean CIELAB male minus mean CIELAB female). For four ethnicities, the skin colours of female subjects had higher lightness value than that of the male subjects. The skin colour of the South Asian and African male subjects was less saturated than that of the equivalent female subjects.

Skin whiteness and blackness indices

The skin colour distribution of each ethnicity L*C*$_{ab}$ plane revealed a trend of skin colour distribution. Adopted the Vividness and Depth formulae (Berns, 2014), a skin whiteness and blackness indices were developed. The skin whiteness index was for Caucasian and Oriental groups. The skin blackness index was for African groups. Different ethnic groups have different skin whiteness /blackness coefficients. The skin whiteness and blackness indices have good linear relationship with the Individual Typology Angle (ITA˚) which has been used in the cosmetics industry to describe the skin whiteness via CIELAB L^* and b^*.

6.1.2 The findings from Experiment 2 psychophysical experiment

The impression and skin appearance attribute that studies previously were selected to investigate in the present study. Twenty-three attributes were selected, including *Dull-Lively*, *Feminine-Masculine*, *Unnatural-Natural, Stupid-Intelligent*, *Babylike-Mature*, *Dislikeable-Likable*, *Serious-Funny*, *Unhealthy-Healthy*, *Unattractive-Attractive*, *Aged-Youth*, *Fussy-Easygoing*, *Tensed-Relaxed*, *Ordinary-Imaginative*, *Loner-Cooperative*, *Passive-Active*, *Autistic-Sociable*, *Pale-Pink*, *Tan-Fair*, *Opaque-Translucent*, *Dry-Moist*, *Matt-Glossy, Rough-Smooth* and *Blemished-Clear*.

To determine the impact of the skin colour on these attributes, the selected images were rendered into 26 images with the difference in skin colour only. In the present study, four images (an Oriental female, an Oriental male, a Caucasian female and a Caucasian male) were selected to be scales. The skin colour that chosen to examine here were selected in terms of the skin whiteness and hue angle values.

Two experiments (Experiment 2.1 and Experiment 2.2) were carried out to determine the impact of the skin colour on these selected attributes.

Experiment 2.1 was to determine the dimensions of the attributes when they were used to scale the images (an Oriental female and an Oriental male) with different skin colours. Ten Chinese observers participated this experiment. The experimental results of Experiment 2.1 showed that four attributes (Likeable, Feminine, Youth and Sociable) can be used to represent the twenty-three attributes. in the present study, *Health* was also chosen to be further studied.

Experiment 2.2 used four selected images to determine the impact of the skin colour on the five impression attributes (Likeable, Feminine, Youth, Sociable and Health). Twenty-Four Chinese observers participated this experiment. The statistical analysis showed that the skin whiteness had a significant impact on judging all the attributes. The hue angle of the skin colour had a significant impact on judging *Sociable* and *Likeable*. Further analysis was conducted by drawing line charts (the mean visual score from the 4 attributes plotted against whiteness and hue angle respectively). The results were summarised as below:

For the Chinese observer with different genders

- The female and male Chinese observers had good agreement at scale the images with different skin whiteness;
- The female and male Chinese observers had a poor disagreement at scaling the images with some of the skin hue angles (mostly at the hue angle equal to 20˚ and 80˚).

For the impact of the skin whiteness:

- The trend of the judgement results of the Feminine and Youth attributes were similar, for the OF and the CF image.
- For the OF and CF images, the feminine, youthfulness, health, sociable and likeable increased when the skin whiteness increased (skin whiteness from 50 to 80);
- For the OM and CM images, the feminine and youthfulness were increased when the skin whiteness increase (skin whiteness from 50 to 80). The OM and CM images appeared the most sociable, likeable and healthiest with skin whiteness equal to 60 and 70, respectively.

For the impact of the skin hue angle:

- The OF and CF images appeared the most feminine, sociable, youthful, likeable and healthiest when the hue angle of the skin equal to 40˚;
- The OM image appeared to be the most feminine and youthful when the hue angle of the skin equal to 50˚, and appeared to be the most sociable, likeable and healthiest when hue angle of the skin equal to 40˚.
- No clear trend was found in the variation of the judgements when observing the CM images with different skin hue angles.

Models were attempted to develop to simulate the relationship between the colorimetric values (skin whiteness and hue angle) and the perceived facial impression attributes for each ethnic and gender group. The models showed acceptable accuracy.

6.2 Contributions

There are 4 main contributions of the present study to the colour science and industrial applications.

1. A comprehensive skin colour database.

The LLSC database is different from the earlier works. It is unique to include a large number of measurement results (from 4 ethnic groups) from the same location by using different measurement methods. Both facial location and body locations were included in this database. This database can be used for academic research and for industrial applications, such as correlation the measurement results between different instruments; the colour difference between body and facial locations, etc.

The LLSC database was accumulated by closely following the most of the guidelines of CIE TC1-92 *Skin Colour Database, which were "with the goal to investigate the uncertainty in skin colour measurement" and "to accumulate skin colour measurements that accord with these protocols covering different ethnicity, gender, and body location.".* This database also a part of the CIE TC1-92 *Skin Colour Database.*

2. Develop the new skin whiteness and blackness indices.

The present study proposed the skin whiteness and blackness indices to determine the whiteness and blackness of the skin colour, respectively. These indices were developed based on the measurement results of real human skin colour and have a good agreement to the ITA˚ which is used in the cosmetic area for determining the lightness (not the lightness of the CIELAB) of the skin colour. This scale can be used to develop a personalised skin whiteness trend line to simulate the skin whiteness change of a specific person.

3. Model the facial impression and skin colour

The present study investigated the frequently used attributes that related to the facial impressions. The experiment found that the 23 attributes, including the skin appearance and perceived impression attributes, can be represented by *Sociable*, *Likeable*, *Feminine* and *Youth.* The present study also determined the skin colour of the both genders and the two examined ethnic groups that appeared the most sociable, likeable, feminine and youthful.

4. Confirmed the skin colour preference of the Orientals.

The results of the preferred skin colour of the presented study agreed with the studies that carried out by the other researchers in the past. The reddish and whiter skin colour was preferred by the Orientals (Chinese and Japanese).

6.3 Future work

The present study gives a good demonstration of appearance science, i.e. to derive a method to quantify human skin colours from the physical measurement to the psychophysical assessment, and models were derived to describe human skin colours. However, some areas can be strengthened and new applications can be carried out using this database.

The future works and new applications could include the following areas:

- For Experiment 1, current LLSC data did not include sufficient subjects for the African and South Asian groups. This leads to some statistical analysis in these groups may not be significant. More skin colour data of the subjects of South Asian and African groups should be accumulated in the further work.
- For Experiment 2, the same experimental setup can be used to be assessed by the Chinese observers from different ethnic groups. This can reveal the culture difference.

- The experiment with more facial images and more skin colour can be carried out to test and to optimise the models that were built in the present study.
- The facial images accumulated in the present study can be applied for different appearance studies, such as to quantify the colour appearance such as whiteness and blackness which was found in the present study.
- The data in the LLSC database can be used to derive models to transform the measurement results between different instruments. This will allow the skin colours data that were measured via the different types of instruments (TSR and SP) to be unified. The measurement results from the instruments also can be transformed to the visual assessment results and reverse.

List of Reference

Adams, L. W. 2010. Whiteness, chromaticness, and blackness symmetries for CIELAB, with a numerical example. Color Research & Application, 35(4), 319-323.

Abdi, H. and Williams, L.J. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics.* **2**(4), pp.433-459.

Adby, M. 2012. Pressing the Beauty Key. *Household and Personal Care Today* **Vol. 7**, pp.P.P.42-47.

Adoniscik, 2008a, CIE CRI TCS SPDs [ONLINE]. Available at: https://commons.wikimedia.org/wiki/File:CIE_CRI_TCS_SPDs.svg [Accessed March 2017].

Adoniscik. 2008b. *(u', v') chromaticity diagram.* www.wikimedia.org.

Agency, A.R.P.a.N.S. 2017. *Fitzpatrick Skin Type.* [Online]. [Accessed Jan]. Available from: http://www.arpansa.gov.au/pubs/RadiationProtection/FitzpatrickSkinType.pdf

Al-Tairi, Z.H., Rahmat, R.W.O., Saripan, M.I. and Sulaiman, P.S. 2014. Skin Segmentation Using YUV and RGB Color Spaces. *JIPS.* **10**(2), pp.283-299.

Alam, M. 2010. *Cosmetic dermatology for skin of color.* McGraw-Hill.

Alaluf S, Atkins D, Barrett K, Blount M, Carter N, Heath A. 2002. The impact of epidermal melanin on objective measurements of human skin color. *Pigment Cell Res*; 15: 119–126

Aoki, K. 2002. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Annals of human biology.* **29**(6), pp.589-608.

Ashikari, M. 2005. Cultivating Japanese whiteness: The 'whitening'cosmetics boom and the Japanese identity. Journal of Material Culture, 10(1), 73-91.

Bartleson, C.J. 1960. Memory colors of familiar objects. *JOSA.* **50**(1), pp.73-77.

Barlow, H. B. 1972. Dark and light adaptation: psychophysics. In Visual psychophysics Springer, Berlin, Heidelberg. pp. 1-28.

BenRG. 2009. *CIE 1931 xy color space diagram.* www.wikimedia.org.

Berns, R.S. 1996a. Deriving instrumental tolerances from pass - fail and colorimetric data. *Color Research & Application.* **21**(6), pp.459-472.

Berns, R.S. 1996b. Methods for characterizing CRT displays. *Displays.* **16**(4), pp.173-182.

Berns, R.S. 2000. *Billmeyer and Saltzman's principles of color technology.* Wiley New York.

Berns, R.S. 2014. Extending CIELAB: Vividness, Vab*, depth, Dab*, and clarity, Tab. *Color Research & Application.* **39**(4), pp.322-330.

Berns, R.S., Motta, R.J. and Gorzynski, M.E. 1993. CRT colorimetry. Part I: Theory and practice. *Color Research & Application.* **18**(5), pp.299-314.

Berry, D.S. and McArthur, L.Z. 1985. Some components and consequences of a babyface. *Journal of personality and social psychology.* **48**(2), p312.

Billmeyer, F.W. and Alessi, P.J. 1981. Assessment of Color - Measuring Instruments. *Color Research & Application.* **6**(4), pp.195-202.

Billmeyer, F.W. and Bencuya, A.K. 1987. Interrelation of the Natural Color System and the Munsell color order system. *Color Research & Application.* **12**(5), pp.243-255.

Broca, P. 1879. Instructions relatives à l'étude anthropologique du système dentaire. *Bulletins de la Société d'anthropologie de Paris.* **2**(1), pp.128-163.

Brydges, D., Deppner, F., Kuenzli, H., Heuberger, K. and Hersch, R.D. 1998. Application of a 3-CCD color camera for colorimetric and densitometric measurements. In: *Photonics West'98 Electronic Imaging*: International Society for Optics and Photonics, pp.292-301.

Burns, R.B. and Dobson, C.B. 2012. *Experimental psychology: Research methods and statistics.* Springer Science & Business Media.

Carré, J.M. and McCormick, C.M. 2008. In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society of London B: Biological Sciences.* **275**(1651), pp.2651-2656.

Carré, J.M., McCormick, C.M. and Mondloch, C.J. 2009. Facial structure is a reliable cue of aggressive behavior. *Psychological Science.* **20**(10), pp.1194-1198.

Chardon, A., Cretois, I. and Hourseau, C. 1991. Skin colour typology and suntanning pathways. *International journal of cosmetic science.* **13**(4), pp.191-208.

Cheung, T. L. V. 2004. Approaches to colour camera characterization (*Doctoral dissertation, University of Leeds*).

Cheung, T.L.V. and Westland, S. 2002. Color camera characterisation using artificial neural networks. In: *Color and Imaging Conference*: Society for Imaging Science and Technology, pp.117-120.

Cheung, V., Li, C., Hardeberg, J., Connah, D. and Westland, S. 2005. Characterization of trichromatic color cameras by using a new multispectral imaging technique. *JOSA A.* **22**(7), pp.1231-1240.

Cheung, V., Westland, S., Connah, D. and Ripamonti, C. 2004. A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms. *Coloration technology.* **120**(1), pp.19-25.

Cho, Y. J., Ou, L. C., Cui G. and Luo M. R. 2017. New colour appearance scales for describing saturation, vividness, blackness, and whiteness. Color Research & Application.

CIE 2004. *CIE 15:2004 Colorimetry.* Vienna.

CIE 2006. *CIE Standard S 014-2E:2006 Colorimetry - Part 2: CIE Standard illiminants.*

CIE 1986. Colorimetry Second Edition: The Evaluation of Whiteness. *CIE Publication* **15.2**, pp.pp. 36–38.

CIE 1987. *CIE017.4: 1987 International lighting vocabulary.* Vienna, Austria: Commission Internationale de l' Eclairage.

CIE 1995a. *CIE013.3: 1995 Method of measuring and specifying colour rendering properties of light sources.* Vienna, Austria: Commission Internationale de l' Eclairage.

CIE 1995b. *CIE 119:1995 Industrial colour difference evaluation.* Vienna, Austria: Commission Internationale de l' Eclairage.

CIE 2001. *CIE 142:2001 Improvement to industrial colour difference evaluation.* Vienna, Austria Commission Internationale de l' Eclairage.

CIE 2004. *CIE 15:2004 Colorimetry.* Vienna, Austria: Commission Internationale de l' Eclairage.

CIE 2011. *S 017/E:2011 ILV: International Lighting Vocabulary.* Vienna, Austria Commission Internationale de l' Eclairage.

Clarke, F., McDonald, R. and Rigg, B. 1984. Modification to the JPC79 colour–difference formula. *Journal of the Society of Dyers and Colourists.* **100**(4), pp.128-132.

Clonts, R., Shamey, R. and Hinks, D. 2010. Effect of colorimetric attributes on perceived blackness of materials. In: *Conference on Colour in Graphics, Imaging, and Vision*: Society for Imaging Science and Technology, pp.83-87.

Clarke, F. J. J., McDonald, R. and Rigg, B. 1984. Modification to the JPC79 colour–difference formula. Coloration Technology, 100(4), 128-132.

Cooper, F. G. (1941). Munsell manual of color.

Coppel, L., Lindberg, S. and Rydefalk, S. 2007. Whiteness assessment of paper samples at the vicinity of the upper CIE whiteness limit.

Cumming, J. and Wooff, D.A. 2007. Dimension reduction via principal variables. *Computational Statistics & Data Analysis.* **52**(1), pp.550-565.

Cunningham, M.R. 1986. Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of personality and social psychology.* **50**(5), p925.

David, A., Krames, M.R. and Houser, K.W. 2013. Whiteness metric for light sources of arbitrary color temperatures: proposal and application to light-emitting-diodes. *Optics express.* **21**(14), pp.16702-16715.

Davis, W. and Ohno, Y. 2010. Color quality scale. *Optical Engineering.* **49**(3), pp.033602-033602-033616.

De Rigal, J., Abella, M.L., Giron, F., Caisey, L. and Lefebvre, M.A. 2007. Development and validation of a new Skin Color Chart®. *Skin Research and Technology.* **13**(1), pp.101-109.

Del Bino, S. and Bernerd, F. 2013. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *British Journal of Dermatology.* **169**(s3), pp.33-40.

Del Bino, S., Sok, J., Bessac, E. and Bernerd, F. 2006. Relationship between skin response to ultraviolet exposure and skin color type. *Pigment cell research.* **19**(6), pp.606-614.

Dictionary, O. E. 2007. Oxford English dictionary online.

Edwards, E.A. and Duntley, S.Q. 1939. The pigments and color of living human skin. *American Journal of Anatomy.* **65**(1), pp.1-33.

Eizoglobal, 2012. *ColorEdge brochure 2012 March edition.* [Online database].

Etcoff, N.L., Stock, S., Haley, L.E., Vickery, S.A. and House, D.M. 2011. Cosmetics as a feature of the extended human phenotype: Modulation of the perception of biologically important facial signals. *PloS one.* **6**(10), pe25656.

Fairchild, M.D. 2005. Color appearance models. 2nd ed. Chichester: Wiley.

Fairchild, M.D. and Reniff, L., 1995. Time course of chromatic adaptation for color-appearance judgments. JOSA A, 12(5), pp.824-833.

Fan, Y., Deng, P., Tsuruoka, H., Aoki, N. and Kobayashi, H. 2011. On the Preferred Flesh Color of Japanese and Chinese and the Determining FactorsInvestigation of the Younger Generation Using Method of Successive Categories and Semantic Differential Method. Journal of The Society of Photographic Science and Technology of Japan, 73, 31-38.

Fasugba, O., Gardner, A. and Smyth, W. 2014. The Fitzpatrick Skin Type Scale: A reliability and validity study in women undergoing radiation therapy for breast cancer. *Journal of wound care.* **23**(7).

Fink, B., Grammer, K. and Matts, P.J. 2006. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior.* **27**(6), pp.433-442.

Fink, B., Grammer, K. and Thornhill, R. 2001. Human (Homo sapiens) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology.* **115**(1), p92.

Fink, B. and Matts, P. 2008. The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology.* **22**(4), pp.493-498.

Fink, B., Matts, P., D'Emiliano, D., Bunse, L., Weege, B. and Röder, S. 2012. Colour homogeneity and visual perception of age, health and attractiveness of male facial skin. *Journal of the European Academy of Dermatology and Venereology.* **26**(12), pp.1486-1492.

Fink, B., Matts, P.J., Klingenberg, H., Kuntze, S., Weege, B. and Grammer, K. 2008. Visual attention to variation in female facial skin color distribution. *Journal of Cosmetic Dermatology.* **7**(2), pp.155-161.

Finlayson, G.D. and Drew, M.S. 1997. Constrained least-squares regression in color spaces. *Journal of Electronic Imaging.* **6**(4), pp.484-493.

Fitzpatrick, T.B. 1988. The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology.* **124**(6), pp.869-871.

Frost, P. 1994. Preference for darker faces in photographs at different phases of the menstrual cycle: preliminary assessment of evidence for a hormonal relationship. *Perceptual and motor skills.* **79**(1), pp.507-514.

Ganz, E. 1972. Whiteness measurement. *J. Col. Appearance.* **1**.

Ganz, E. and Pauli, H.K.A. 1995. Whiteness and tint formulae of the Commission Internationale de l'Eclairage: approximations in the L* a* b* color space. *Applied optics.* **34**(16), pp.2998-2999.

Georgoula, M. 2015. *Assessing Colour Differences Of Lighting Stimuli Using A Visual Display*. Ph.D. thesis, University of Leeds.

Gescheider, G. A. (2013). Psychophysics: the fundamentals. Psychology Press.

Glassner, A.S. 1989. How to derive a spectrum from an RGB triplet. *IEEE Computer Graphics and Applications.* **9**(4), pp.95-99.

Gloag, H. and Gold, M. 1978. Colour Co-ordination Handbook. *London: Her Majesty's Stationary Office.*

Gong, R., Xu, H. and Tong, Q. 2012. Colorimetric characterization models based on colorimetric characteristics evaluation for active matrix organic light emitting diode panels. *Applied optics.* **51**(30), pp.7255-7261.

Grammer, K. and Thornhill, R. 1994. Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of comparative psychology.* **108**(3), p233.

Green, P. 1999. *Understanding digital color.* GATFPress.

Green, P., 2002, "6 Overview of characterization methods- Colour engineering: achieving device independent colour by Green, P., & MacDonald, L. (2002). John Wiley & Sons.

Grimm, K.A., Foltz, J.L., Blanck, H.M. and Scanlon, K.S. 2012. Household income disparities in fruit and vegetable consumption by state and territory: results of the 2009 Behavioral Risk Factor Surveillance System. *Journal of the Academy of Nutrition and Dietetics.* **112**(12), pp.2014-2021.

Gunn, D.A., Rexbye, H., Griffiths, C.E., Murray, P.G., Fereday, A., Catt, S.D., Tomlin, C.C., Strongitharm, B.H., Perrett, D.I. and Catt, M. 2009. Why some women look young for their age. *PloS one.* **4**(12), pe8021.

Hainich, R.R. and Bimber, O. 2016. *Displays: fundamentals & applications.* CRC press.

Hallgren, K. A. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials in quantitative methods for psychology, 8(1), 23.

Hardeberg, J.Y. 2001. Acquisition and reproduction of color images*: colorimetric and multispectral approaches.* Universal-Publishers.

Haslup, J., Shamey, R. and Hinks, D. 2013. The effect of hue on the perception of blackness using munsell samples. *Color Research & Application.* **38**(6), pp.423-428.

He, G. and Zhang, Z. 2002. New whiteness formula in the CIELUV uniform color space. In: *Photonics Asia 2002*: International Society for Optics and Photonics, pp.163-166.

He, G. and Zhou, M. 2007. Whiteness formula in CIELAB uniform color space. *Chinese Optics Letters.* **5**(7), pp.432-434.

Hehman, E., Leitner, J.B., Deegan, M.P. and Gaertner, S.L. 2013a. Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological science.* **24**(3), pp.289-296.

Hehman, E., Leitner, J.B. and Gaertner, S.L. 2013b. Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology.* **49**(4), pp.747-754.

Hill, M.E. 2002. Skin color and the perception of attractiveness among African Americans: Does gender make a difference? *Social Psychology Quarterly.* pp.77-91.

Hong, G., Luo, M.R. and Rhodes, P.A. 2001. A study of digital camera colorimetric characterisation based on polynomial modelling.

Howitt, D. and Cramer, D. 2011. *Introduction to SPSS statistics in psychology: for version 19 and earlier.* Pearson.

Hung, P.-C. 1991. Colorimetric calibration for scanners and media. In: *Electronic Imaging'91, San Jose, CA*: International Society for Optics and Photonics, pp.164-174.

Hunt, R.W.G., 1950. The effects of daylight and tungsten light-adaptation on color perception. JOSA, 40(6), pp.362-371.

Hunt, R. 1986. The Colour Science of Dyes and Pigments. *Journal of Modern Optics.* **33**(8), pp.949-950.

Hunt, R.W.G. and Pointer, M.R. 2011. *Measuring colour.* John Wiley & Sons.

Hunter, R.S. and Harold, R.W. 1987. *The measurement of appearance.* John Wiley & Sons.

ISO 2003. *ISO/TR 16066:2003 Graphic technology -- Standard object colour spectra database for colour reproduction evaluation (SOCS).* Switzerland: International Organization for Standardization.

ISO 2007. ISO 11664-2:2007 Colorimetry — Part 2: CIE standard illuminants.

Jablonski, N.G. 2004. The evolution of human skin and skin color. *Annu. Rev. Anthropol.* **33**, pp.585-623.

Jablonski, N.G. and Chaplin, G. 2000. The evolution of human skin coloration. *Journal of human evolution.* **39**(1), pp.57-106.

Jacobolus. 2007. Munsell color system. www.wikimedia.org: wikimedia.

Jafari, R., Amirshahi, S. and Ravandi, S.H. 2016. Colorimetric analysis of black coated fabrics. *Journal of Coatings Technology and Research.* **13**(5), pp.871-882.

Jafari, R. and Amirshahi, S.H. 2007. A comparison of the CIE and Uchida whiteness formulae as predictor of average visual whiteness evaluation of textiles. *Textile Research Journal.* **77**(10), pp.756-763.

Jameson, D., Hurvich, L.M. and Varner, F.D., 1979. Receptoral and postreceptoral visual processes in recovery from chromatic adaptation. Proceedings of the National Academy of Sciences, 76(6), pp.3034-3038.

Jiang, Z. X. and Kaplan, P. D. 2008. Point‐spread imaging for measurement of skin translucency and scattering. Skin Research and Technology, 14(3), 293-297.

Jiang, Z. X. and DeLaCruz, J. 2011. Appearance benefits of skin moisturization. Skin research and technology, 17(1), 51-55.

John, O.P. and Srivastava, S. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research.* **2**(1999), pp.102-138.

Joiner, A., Hopkinson, I., Deng, Y. and Westland, S. 2008. A review of tooth colour and whiteness. *Journal of dentistry.* **36**, pp.2-7.

Jolliffe, I. 2002. *Principal component analysis.* Wiley Online Library.

Jones, B., Little, A., Feinberg, D., Penton-Voak, I., Tiddeman, B. and Perrett, D. 2004. The relationship between shape symmetry and perceived skin condition in male facial attractiveness. *Evolution and Human Behavior.* **25**(1), pp.24-30.

Jones, B.C., Little, A.C., Burt, D.M. and Perrett, D.I. 2004. When facial attractiveness is only skin deep. *Perception.* **33**(5), pp.569-576.

Jones, B.C., Little, A.C., Penton-Voak, I.S., Tiddeman, B., Burt, D.M. and Perrett, D. 2001. Facial symmetry and judgements of apparent health: support for a “good genes” explanation of the attractiveness–symmetry relationship. *Evolution and human behavior.* **22**(6), pp.417-429.

Jones, M.J. and Rehg, J.M. 2002. Statistical color models with application to skin detection. *International Journal of Computer Vision.* **46**(1), pp.81-96.

Jordan, B. and O'neill, M. 1991. The whiteness of paper-colorimetry and visual ranking. *Tappi journal.* **74**(5), pp.93-101.

Katayama, I. and Fairchild, M.D. 2010. Quantitative evaluation of perceived whiteness based on a color vision model. *Color Research & Application.* **35**(6), pp.410-418.

Konica Minolta. 2017. Luminance vs. Illuminance. [ONLINE] Available at: http://sensing.konicaminolta.us/2015/08/luminance-vs-illuminance/. [Accessed July 2017].

Kuehni, R.G. and Schwarz, A. 2008. *Color ordered: a survey of color systems from antiquity to the present.* Oxford University Press.

Lefevre, C.E. and Perrett, D.I. 2015. Fruit over sunbed: Carotenoid skin colouration is found more attractive than melanin colouration. *The Quarterly Journal of Experimental Psychology.* **68**(2), pp.284-293.

Lévêque, J.-L. and Goubanova, E. 2004. Influence of age on the lips and perioral skin. *Dermatology.* **208**(4), pp.307-313.

Li, C., Cui, G. and Luo, M.R. 2003. The accuracy of polynomial models for characterising digital cameras. *Proc. AIC Interim. Meet.* pp.166-170.

Li, E.P., Min, H.J. and Belk, R.W. 2008a. Skin lightening and beauty in four Asian cultures. *NA-Advances in Consumer Research Volume 35.*

Li, E.P., Min, H.J. and Belk, R.W. 2008b. Skin lightening and beauty in four Asian cultures. *ACR North American Advances.*

Lin, J., Shamey, R. and Hinks, D. 2012. Factors affecting the whiteness of optically brightened material. *JOSA A.* **29**(11), pp.2289-2299.

Little, A.C., Burt, D., Penton-Voak, I. and Perrett, D. 2001. Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society of London B:* Biological Sciences. 268(1462), pp.39-44.

Luo, M. R., Cui, G. and Rigg, B. 2001. The development of the CIE 2000 colour‐difference formula: CIEDE2000. Color Research & Application, 26(5), 340-350.

Luo, M. R. and Rhodes, P., 1999, Lecture 4 Colour measuring instruments, *MSc colour and imaging science, University of Leeds.*

Luo, W., Westland, S., Brunton, P., Ellwood, R., Pretty, I.A. and Mohan, N. 2007. Comparison of the ability of different colour indices to assess changes in tooth whiteness. *journal of dentistry.* **35**(2), pp.109-116.

Luo, W., Westland, S., Ellwood, R. and Pretty, I. 2005. Evaluation of whiteness formulae for teeth. In: *Proceedings of the 10th congress of the international colour association*: Graficas Alhambra, Granada, pp.100-104.

Luo, W., Westland, S., Ellwood, R., Pretty, I. and Cheung, V. 2009. Development of a whiteness index for dentistry. *Journal of dentistry.* **37**, pp.e21-e26.

Magin, P., Pond, D., Smith, W., Goode, S. and Paterson, N. 2012. Reliability of skin‐type self‐assessment: agreement of adolescents' repeated Fitzpatrick skin phototype classification ratings during a cohort study. *Journal of the European Academy of Dermatology and Venereology.* **26**(11), pp.1396-1399.

Malacara, D. 2003. *Color vision and colorimetry: theory and applications*. Wiley Online Library.

Mantiuk, R.K., Tomaszewska, A. and Mantiuk, R. 2012. Comparison of four subjective methods for image quality assessment. In: *Computer Graphics Forum*: Wiley Online Library, pp.2478-2491.

Mattos, K.P., Lloret, G.R., Cintra, M.L., Gouvêa, I.R., Betoni, T.R., Mazzola, P.G. and Moriel, P. 2016. Acquired skin hyperpigmentation following intravenous polymyxin B treatment: a cohort study. *Pigment cell & melanoma research.* **29**(3), pp.388-390.

Marschner, S.R., Westin, S.H., Lafortune, E.P., Torrance, K.E. and Greenberg, D.P., 1999. Image-based BRDF measurement including human skin. In Rendering Techniques' 99 (pp. 131-144). Springer, Vienna.

Maydeu-Olivares, A., 2004. Thurstone's case V model: A structural equations modeling perspective. In Recent developments on structural equation models (pp. 41-67). Springer Netherlands.

McCrae, R.R. and Costa Jr, P.T. 1999. A five-factor theory of personality. Handbook of personality: Theory and research. 2, pp.139-153.

McDonald, R. and Smith, K. J. 1995. CIE94 - a new colour - difference formula. Coloration Technology, 111(12), 376-379.

McLaren, K. 1970. Colour Passing - Visual or Instrumental? Coloration Technology. 86(9), pp.389-392.

Mealey, L., Bridgstock, R. and Townsend, G.C. 1999. Symmetry and perceived facial attractiveness: a monozygotic co-twin comparison. *Journal of personality and social psychology.* **76**(1), p151.

Middaugh, A.L., Fisk, P.S., Brunt, A. and Rhee, Y.S. 2012. Few associations between income and fruit and vegetable consumption. *Journal of nutrition education and behavior.* **44**(3), pp.196-203.

Morgan, K.V., Morton, A., Whitehead, R.D., Perrett, D.I., Hurly, T.A. and Healy, S.D. 2016. Assessment of health in human faces is context-dependent. *Behavioural processes.* **125**, pp.89-95.

Motokawa, T., Kato, T., Hashimoto, Y., Hongo, M., Ito, M., Takimoto, H. and Katagiri, T. 2006. Characteristic MC1R polymorphism in the Japanese population. *Journal of dermatological science.* **41**(2), pp.143-145.

Muehlenbein, M.P. 2010. *Human evolutionary biology.* Cambridge University Press.

Murakami, Y., Obi, T., Yamaguchi, M., Ohyama, N. and Komiya, Y. 2001. Spectral reflectance estimation from multi-band image using color chart. *Optics communications.* 188(1), pp.47-54.

Munsell, A. H. 1915. Atlas of the Munsell color system. Wadsworth, Howland & Company, Incorporated, Printers.

Nayatani, Y. and Sakai, H. 2008. An integrated color - appearance model using CIELUV and its applications. *Color Research & Application.* **33**(2), pp.125-134.

NCS. 2017. *NCS colour triangle.* http://www.ncscolour.co.uk/information/ncs-system.html: NCS.

Nobbs, J. 2002. A lightness, chroma and hue splitting approach to CIEDE2000 colour differences. *Advances in Colour Science and Technology.* **5**(2), pp.46-53.

Ohta, N. and Robertson, A. 2006. *Colorimetry: fundamentals and applications.* John Wiley & Sons.

Ou, L.C., Luo, M.R., Woodcock, A. and Wright, A. 2004. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application.* **29**(3), pp.232-240.

Parra, E.J. 2007. Human pigmentation variation: evolution, genetic basis, and implications for public health. *American journal of physical anthropology.* **134**(S45), pp.85-105.

Paunonen, S.V., Ewan, K., Earthy, J., Lefave, S. and Goldberg, H. 1999. Facial features as personality cues. *Journal of personality.* **67**(3), pp.555-583.

Penton-Voak, I., Jones, B.C., Little, A., Baker, S., Tiddeman, B., Burt, D. and Perrett, D. 2001. Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London B: Biological Sciences.* **268**(1476), pp.1617-1623.

Perrett, D.I., Burt, D.M., Penton-Voak, I.S., Lee, K.J., Rowland, D.A. and Edwards, R. 1999. Symmetry and human facial attractiveness. *Evolution and human behavior.* **20**(5), pp.295-307.

Pervin, L.A. and John, O.P. 1999. *Handbook of personality: Theory and research.* Elsevier.

Pelli, D.G. and Farell, B., 1995. Psychophysical methods. Handbook of Optics. Second ed. New York: McGraw-Hill.

Picton, O. 2013. The complexities of complexion: a cultural geography of skin colour and beauty products. *Geography.* **98**, p85.

Pointer, M., Attridge, G. and Jacobson, R.E. 2001. Practical camera characterization for colour measurement. *The Imaging Science Journal.* **49**(2), pp.63-80.

Post, D.L. and Calhoun, C.S. 1989. An evaluation of methods for producing desired colors on CRT monitors. *Color Research & Application.* **14**(4), pp.172-186.

Prins, N. 2016. *Psychophysics: a practical introduction.* Academic Press.

Publ., C. 2005. *167:2005. Recommended practice for tabulating spectral data for use in colour computations.*

Rhodes, A.,P. 2002, “5 colour notaion system - Colour engineering: achieving device independent colour by Green, P. and MacDonald, L. (2002). John Wiley & Sons.

Rhodes, G., Proffitt, F., Grady, J.M. and Sumich, A. 1998. Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review.* **5**(4), pp.659-669.

Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L.W., Peters, M., Lee, K., Halberstadt, J. and Crawford, J.R. 2007. Perceived health contributes to the attractiveness of facial symmetry, averageness, and sexual dimorphism. *Perception.* **36**(8), pp.1244-1252.

Rhodes, G., Zebrowitz, L.A., Clark, A., Kalick, S.M., Hightower, A. and McKay, R. 2001. Do facial averageness and symmetry signal health? *Evolution and Human Behavior.* **22**(1), pp.31-46.

Robins, A.H. 2005. *Biological perspectives on human pigmentation.* Cambridge University Press.

Samson, N., Fink, B., Matts, P.J., Dawes, N.C. and Weitz, S. 2010. Visible changes of female facial skin surface topography in relation to age and attractiveness perception. *Journal of cosmetic dermatology.* **9**(2), pp.79-88.

Sander, E., Warner, R., Harrison, H. and Loosli, J. 1959. The stimulatory effect of sodium butyrate and sodium propionate on the development of rumen mucosa in the young calf. *Journal of Dairy Science.* **42**(9), pp.1600-1605.

Sani, F. and Todman, J. 2008. *Experimental design and statistics for psychology: a first course.* John Wiley & Sons.

Scheib, J.E., Gangestad, S.W. and Thornhill, R. 1999. Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences.* **266**(1431), pp.1913-1917.

Sephora. 2017. Sephora+Pantone color IQ. [ONLINE]. Available from: http://www.sephora.com/color-iq [Accessed Jan 2017].

Sersen, W. 1989. Buying and Selling Gems: What Light is Best. What lighting do you use when buying stones, 13-23.

Sharma, G. and Bala, R. 2002. *Digital color imaging handbook.* CRC press.

Shen, H.-L., Zhang, H.-G., Xin, J.H. and Shao, S.-J. 2008. Optimal selection of representative colors for spectral reflectance reconstruction in a multispectral imaging system. *Applied optics.* **47**(13), pp.2494-2502.

Shimano, N., Terai, K. and Hironaga, M. 2007. Recovery of spectral reflectances of objects being imaged by multispectral cameras. *JOSA A.* **24**(10), pp.3211-3219.

Shriver MD, Parra EJ. 2000, Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. *Am J Phys Anthropol*; 112: 17–27.

Soriano, M., Marszalec, E. and Pietikainen, M. 2000. Physics-based face database for color research. *Journal of Electronic Imaging.* **9**(1), pp.32-38.

Soriano, M., Melgosa, M., Sánchez-Marañón, M., Delgado, G., Gámiz, E. and Delgado, R. 1998. Whiteness of talcum powders as a quality index for pharmaceutical uses. *Color research and application.* **23**(3), pp.178-185.

Stamatas, G.N., Zmudzka, B.Z., Kollias, N. and Beer, J.Z. 2004. Non‐invasive measurements of skin pigmentation in situ. *Pigment Cell Research.* **17**(6), pp.618-626.

Standard, B. 1988. 6923: 1988, British Standard method for calculation of small colour differ-ences. *British Standards Institution, Milton Keynes.*

Stephen, I.D., Coetzee, V. and Perrett, D.I. 2011. Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behavior.* **32**(3), pp.216-227.

Stephen, I.D., Coetzee, V., Smith, M.L. and Perrett, D.I. 2009a. Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS One.* **4**(4), pe5083. Change skin colour for healthy look

Stephen, I.D. and McKeegan, A.M. 2010. Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception.* **39**(8), pp.1104-1110.

Stephen, I.D., Scott, I.M., Coetzee, V., Pound, N., Perrett, D.I. and Penton-Voak, I.S. 2012. Cross-cultural effects of color, but not morphological masculinity, on perceived attractiveness of men's faces. *Evolution and Human Behavior.* **33**(4), pp.260-267.

Stephen, I.D., Smith, M.J.L., Stirrat, M.R. and Perrett, D.I. 2009b. Facial skin coloration affects perceived health of human faces. *International journal of primatology.* **30**(6), pp.845-857.

Stirrat, M. and Perrett, D.I. 2010. Valid facial cues to cooperation and trust male facial width and trustworthiness. *Psychological science.* **21**(3), pp.349-354.

Stirrat, M. and Perrett, D.I. 2012. Face Structure Predicts Cooperation Men With Wider Faces Are More Generous to Their In-Group When Out-Group Competition Is Salient. *Psychological science.* p0956797611435133.

Sun, Q. and Fairchild, M.D. 2002. Statistical characterization of face spectral reflectances and its application to human portraiture spectral estimation. *Journal of Imaging Science and Technology.* **46**(6), pp.498-506.

Swami, V., Furnham, A. and Joshi, K. 2008. The influence of skin tone, hair length, and hair colour on ratings of women's physical attractiveness, health and fertility. *Scandinavian Journal of Psychology.* **49**(5), pp.429-437.

Swiatoniowski, A.K., Quillen, E.E., Shriver, M.D. and Jablonski, N.G. 2013. Technical note: Comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation. *American journal of physical anthropology.* **151**(2), pp.325-330.

Tajima, J., Tsukada, M., Miyake, Y., Haneishi, H., Tsumura, N., Nakajima, M., Azuma, Y., Iga, T., Inui, M. and Ohta, N. 1998. Development and standardization of a spectral characteristics database for evaluating color reproduction in image input devices. In: *SYBEN-Broadband European Networks and Electronic Image Capture and Publishing*: International Society for Optics and Photonics, pp.42-50.

Takiwaki H, Ovengaard L, Serup J. 1994. Comparison of narrow-band reflectance spectrophotometric and tristimulus colorimetric measurements of skin color. *Skin Pharmacol*; 7: 217–225

Takiwaki H, Miyaoka Y, Kohno H, Arase S. 2002. Graphic analysis of the relationship between skin colour change and variations in the amounts of melanin and haemoglobin. *Skin Res Tech*; 8: 78–83.

Tao, L., Westland, S. and Cheung, V. 2011. Blackness: preference and perception. In: *Color and Imaging Conference*: Society for Imaging Science and Technology, pp.270-275.

Taylor, S., Westerhof, W., Im, S. and Lim, J. 2006. Noninvasive techniques for the evaluation of skin color. *Journal of the American Academy of Dermatology.* **54**(5), pp.S282-S290.

Taylor, S.C., Arsonnaud, S. and Czernielewski, J. 2005. The Taylor Hyperpigmentation Scale: a new visual assessment tool for the evaluation of skin color and pigmentation. *Cutis.* **76**(4), pp.270-274.

Terapeak. 2014. *Photo Research Pr-650 Spectrascan Colorimeter W/ Pr Ms-75 Macro.*

Thomas, J.-B. and Hardeberg, J.Y. 2013. Cross-Media Color Reproduction and Display Characterization. In: Fernandez-Maloigne, C. ed. *Advanced Color Image Processing and Analysis.* New York: Springer

Thomas, J.B., Hardeberg, J.Y., Foucherot, I. and Gouton, P. 2008. The PLVC display color characterization model revisited. *Color Research and Application.* **33**(6), pp.449-460.

Tinsley, H.E. and Brown, S.D. 2000. *Handbook of applied multivariate statistics and mathematical modeling.* Academic Press.

Treesirichod, A., Chansakulporn, S. and Wattanapan, P. 2014. Correlation between skin color evaluation by skin color scale chart and narrowband reflectance spectrophotometer. Indian journal of dermatology. 59(4), p339.

Tsukida, K. and Gupta, M. R. 2011. How to analyze paired comparison data (No. UWEETR-2011-0004). WASHINGTON UNIV SEATTLE DEPT OF ELECTRICAL ENGINEERING.

Tsumura, N., Usuba, R., Takase, K., Nakaguchi, T., Ojima, N., Komeda, N., and Miyake, Y. 2008. Image-based control of skin translucency. Applied optics, 47(35), 6543-6549.

Tsumura, N., Kawabuchi, M., Haneishi, H. and Miyake, Y., 2001. Mapping pigmentation in human skin from a multi-channel visible spectrum image by inverse optical scattering technique. Journal of Imaging Science and Technology, 45(5), pp.444-450.

Tsumura, N., Ojima, N., Sato, K., Shiraishi, M., Shimizu, H., Nabeshima, H., Akazaki, S., Hori, K. and Miyake, Y., 2003. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. ACM Transactions on Graphics (TOG), 22(3), pp.770-779.

Uchida, H. 1990. A study on the CIE whiteness formula. *J. Color Sci. Asso. Jpn.* **14**, pp.106-113.

Uchida, H. 1998. A new whiteness formula. *Color Research and Application.* **23**(4), pp.202-209.

Van den Berghe, P.L. and Frost, P. 1986. Skin color preference, sexual dimorphism and sexual selection: A case of gene culture co‐evolution? *Ethnic and Racial Studies.* **9**(1), pp.87-113.

Viera, A. J., and Garrett, J. M. 2005 Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

Wang, H., Cui, G., Luo, M.R. and Xu, H. 2012. Evaluation of colour‐difference formulae for different colour‐difference magnitudes. *Color Research and Application.* **37**(5), pp.316-325.

Wang, Y.Z., Luo, M.R., Liu, X.Y., Qiang, J. and Bian, J. 2016. Skin Lighting Study Based on Multispectral Images. CIE conference proceeding 2016. pp. 446-450

Watier, N., Lamontagne, C. and Chartier, S. 2014. Descriptive Statistics. *Probability and Statistics.* CRC Press, pp.1-37.

Westland, S. 1998. Artificial neural networks explained‐Part 1. *Coloration Technology.* **114**(10), pp.274-276.

Westland, S. 2015. CIE Whiteness.

Westland, S., Cheung, T.L.V. and Lozman, O. 2006. A metric for predicting perceptual blackness. In: *Color and Imaging Conference*: Society for Imaging Science and Technology, pp.14-17.

Whitehead, R.D., Perrett, D.I. and Ozakinci, G. 2012. Attractive skin coloration: Harnessing sexual selection to improve diet and health. *Evolutionary Psychology.* **10**(5), p147470491201000507.

Whitfield, T., O'Connor, M. and Wiltshire, T. 1986. Architects' attitudes to the British building colour standards and colour use in buildings. *Notes.* **181**, pp.181-195.

Wong, E.M., Haselhuhn, M.P. and Kray, L.J. 2012. Improving the future by considering the past: The impact of upward counterfactual reflection and implicit beliefs on negotiation performance. *Journal of Experimental Social Psychology.* **48**(1), pp.403-406.

Xiao, K., Liao, N., Zardawi, F., Liu, H., Van Noort, R., Yang, Z., Huang, M. and Yates, J.M. 2012. Investigation of Chinese skin colour and appearance for skin colour reproduction. *Chinese Optics Letters.* **10**(8), p083301.

Xiao, K., Zhu, Y., Li, C., Connah, D., Yates, J.M. and Wuerger, S. 2016. Improved method for skin reflectance reconstruction from camera images. *Optics express.* **24**(13), pp.14934-14950.

Xiao, K., Yates, J. M., Zardawi, F., Sueeprasan, S., Liao, N., Gill, L., Li, C. and Wuerger, S. 2017. Characterising the variations in ethnic skin colours: a new calibrated data base for human skin. *Skin Research and Technology*, 23(1), 21-29.

Xin, J.H. 2006. Total colour management in textiles. Woodhead Publishing.

Yamamoto, M., Lim, Y. H., Aoki, N., and Kobayashi, H. 2002. On the preferred reproduction of flesh color in Japan and South Korea, investigated for all generation. *Journal of The Society of Photographic Science and Technology of Japan*, 65(5), pp.363-368.

Yano, T. and Hashimoto, K. 1997. Preference for Japanese complexion color under illumination. *Color Research and Application.* **22**(4), pp.269-274.

Yoshikawa, H., Kikuchi, K., Yaguchi, H., Mizokami, Y. and Takata, S. 2012. Effect of chromatic components on facial skin whiteness. *Color Research and Application.* **37**(4), pp.281-291.

Young, T.Y. and Liu, P. 1986. Handbook of pattern recognition and image processing. In: *Academic Press*.

Yuan, Y., Ou, L.-C. and Luo, M.R. 2011. Effects of Skin Tone and Facial Characteristics on Perceived Attractiveness: A Cross-Cultural Study. In: *Color and Imaging Conference*: *Society for Imaging Science and Technology*, pp.253-259.

Zardawi, F. M., Xiao, K. and Yates, J. M. 2015. Skin Colour Analysis of Iraqi Kurdish Population*. Journal of Oral and Dental Research*, 2(1), 4-9.

Zeng, H. and Luo, R. 2009. Modelling skin colours for preferred colour reproduction. In: *Color and Imaging Conference*: Society for Imaging Science and Technology, pp.175-180.

Zeng, H. and Luo, R. 2013. Colour and tolerance of preferred skin colours on digital photographic images. *Color Research and Application.* **38**(1), pp.30-45.

Zhdanova, E.Y., Chubarova, N.Y. and Blumthaler, M. 2014. Biologically active UV-radiation and UV-resources in Moscow (1999–2013). *Geogr. Environ. Sustainability.* **2**, pp.71-85.

List of Abbreviations

AF	African Female
AM	African Male
ANOVA	Analysis of Variance
BH	Back of Hand
CAM	Colour Appearance Model
CB	Cheek Bone
CCT	Correlated Colour Temperature
CF	Caucasian Female
CGFI	Computer Generated Facial Image
CH	Cheek
CIE	Commission Internationale de L' Éclairage (International Commission on Illumination)
CIELAB	CIE 1976 ($L^*a^*b^*$) Colour Space
CIELUV	CIE 1976 ($L^*u^*v^*$) Colour Space
CIEWI	CIE Whiteness Index
CIEXYZ	The CIE Tristimulus Values
CM	Caucasian Male
CM700d	KONICA MINOLTA CM-700d
CRT	Cathode Ray Tube
CS	Consecutive Repeatability Measurement
CT	Continuous Repeatability Measurement
F	Female
FH	Forehead
FT	Finger Tip

GOG	Gamma Offset Gain
HP	High Pressure
IF	Inner Forearm
ISO	International Organization for Standardization
ITA	The Individual Typologic Angle
LCD	Liquid Crystal Display
LED	Light-Emitting Diode
LLSC	Leeds Liverpool Skin Colour Database
LP	Low Pressure
LUT	Lookup Table
M	Male
MAV	Large Aperture Size
MCCC	Gretagmacbeth™ Macbeth Color Checker Chart (24 colours)
MCDM	Mean of Colour Difference from The Mean
mins	minutes
NCS	Natural Color System
NT	Nose Tip
OF	Oriental Female
OF	Outer Forearm
OM	Oriental Male
PCA	Principal Component Analysis
PLCC	The Piecewise Linear Assuming Chromaticity Constancy Model
PR650	Photo Research Spectrascan PR650

RGB	R: Red Channel; G: Green Channel; B: Blue Channel. The Range Of The Value Of The R,G,B Is From 0 To 255
SAF	South Asian Female
SAM	South Asian Male
SAV	The Small Aperture Size
SEM	Standard Error of Mean
SP	Spectrophotometer
SPD	Spectral Power Distribution Or S(Λ)
TSR	Tele-Spectroradiometer
UV	Ultra -Violet
VA	Visual Assessment
WD	Wrong Decision (%)

Appendix A

Monitor gamma evaluation

Table A1 The Channel R raw data of input and measured output (by using the Jeti).

Channel type +test sample number	input			Output (measured)		
	R	G	B	X	Y	Z
R 1	0	0	0	0.203	0.221	0.319
R 2	15	0	0	0.334	0.278	0.338
R 3	30	0	0	0.771	0.486	0.330
R 4	45	0	0	1.460	0.797	0.333
R 5	60	0	0	2.551	1.282	0.350
R 6	75	0	0	4.033	1.939	0.393
R 7	90	0	0	5.958	2.805	0.392
R 8	105	0	0	8.322	3.891	0.461
R 9	120	0	0	11.110	5.134	0.452
R 10	135	0	0	14.364	6.611	0.538
R 11	150	0	0	18.038	8.257	0.636
R 12	165	0	0	22.086	10.060	0.676
R 13	180	0	0	26.622	12.100	0.782
R 14	195	0	0	31.773	14.420	0.808
R 15	210	0	0	37.421	17.000	0.936
R 16	225	0	0	43.914	19.900	1.071
R 17	240	0	0	50.020	22.680	1.152
R 18	255	0	0	57.152	25.880	1.213

Table A2 The Channel G raw data of input and measured output (by using Jeti).

Channel type +test sample number	input			Output (measured)		
	R	**G**	B	X	Y	Z
G 1	0	**0**	0	0.216	0.230	0.299
G 2	0	**15**	0	0.265	0.388	0.348
G 3	0	**30**	0	0.418	0.905	0.414
G 4	0	**45**	0	0.679	1.741	0.503
G 5	0	**60**	0	1.055	3.048	0.706
G 6	0	**75**	0	1.600	4.827	0.943
G 7	0	**90**	0	2.309	7.155	1.265
G 8	0	**105**	0	3.178	10.020	1.651
G 9	0	**120**	0	4.200	13.410	2.117
G 10	0	**135**	0	5.430	17.390	2.674
G 11	0	**150**	0	6.761	21.860	3.269
G 12	0	**165**	0	8.225	26.650	3.940
G 13	0	**180**	0	9.900	32.210	4.653
G 14	0	**195**	0	11.820	38.370	5.540
G 15	0	**210**	0	13.836	45.060	6.399
G 16	0	**225**	0	15.943	51.930	7.366
G 17	0	**240**	0	18.292	59.740	8.497
G 18	0	**255**	0	20.951	68.400	9.708

Table A3 The Channel B raw data of input and measured output (by using Jeti).

Channel type +test sample number	Input			Output (measured)		
	R	G	**B**	X	Y	Z
B 1	0	0	**0**	0.202	0.222	0.299
B 2	0	0	**15**	0.255	0.241	0.537
B 3	0	0	**30**	0.388	0.279	1.275
B 4	0	0	**45**	0.621	0.372	2.487
B 5	0	0	**60**	0.996	0.498	4.416
B 6	0	0	**75**	1.455	0.627	7.006
B 7	0	0	**90**	2.097	0.853	10.362
B 8	0	0	**105**	2.839	1.090	14.421
B 9	0	0	**120**	3.780	1.429	19.260
B 10	0	0	**135**	4.845	1.765	25.078
B 11	0	0	**150**	6.045	2.167	31.477
B 12	0	0	**165**	7.370	2.596	38.648
B 13	0	0	**180**	8.905	3.127	46.746
B 14	0	0	**195**	10.618	3.708	55.901
B 15	0	0	**210**	12.540	4.324	66.130
B 16	0	0	**225**	14.365	4.941	76.080
B 17	0	0	**240**	16.548	5.677	87.582
B 18	**0**	0	**255**	18.748	6.448	99.043

Table A4 The normalised data of RGB and normalised data of Y with light leak at input (0,0,0) removed.

NO.	Input of R/G/B (normalised)	Output R (Y)	Output G (Y)	Output B (Y)
1	0.000	0.000	0.000	0.000
2	0.059	0.002	0.002	0.003
3	0.118	0.010	0.010	0.009
4	0.176	0.022	0.022	0.024
5	0.235	0.041	0.041	0.044
6	0.294	0.067	0.067	0.065
7	0.353	0.101	0.102	0.101
8	0.412	0.143	0.144	0.139
9	0.471	0.191	0.193	0.194
10	0.529	0.249	0.252	0.248
11	0.588	0.313	0.317	0.312
12	0.647	0.383	0.388	0.381
13	0.706	0.463	0.469	0.467
14	0.765	0.553	0.559	0.560
15	0.824	0.654	0.658	0.659
16	0.882	0.767	0.758	0.758
17	0.941	0.875	0.873	0.876
18	1.000	1.000	1.000	1.000

Table A5 The gamma of each channel (R, G and B).

NO.	γ of R channel	γ of G channel	γ of B channel
1	NA	NA	NA
2	2.158	2.142	2.037
3	2.137	2.157	2.190
4	2.189	2.196	2.149
5	2.202	2.202	2.154
6	2.210	2.203	2.233
7	2.204	2.196	2.197
8	2.192	2.187	2.220
9	2.193	2.180	2.176
10	2.186	2.169	2.193
11	2.188	2.163	2.193
12	2.202	2.177	2.215
13	2.211	2.173	2.189
14	2.206	2.165	2.162
15	2.188	2.159	2.149
16	2.120	2.209	2.214
17	2.197	2.241	2.181
18	NA	NA	NA
Average	2.186	2.182	2.178

Appendix B

The range of the skin colours RGB values

Table B1 The range of the RGB values of Oriental and Caucasian.

		R	G	B
Oriental	**Max**	192.90	167.45	149.10
	Min	96.39	73.80	52.74
	Mean	162.67	134.46	111.24
Caucasian	**Max**	194.16	168.22	150.03
	Min	117.78	93.47	69.94
	Mean	175.30	147.84	128.86

The table above listed the range of the mean RGB values of four facial locations of the Oriental and Caucasian group. The four facial locations were forehead, cheekbone, cheek and neck. These four locations were marked and selected by using MATLAB. The average RGB values of each location were used as the RGB value of this location.

The structure of the image processing code

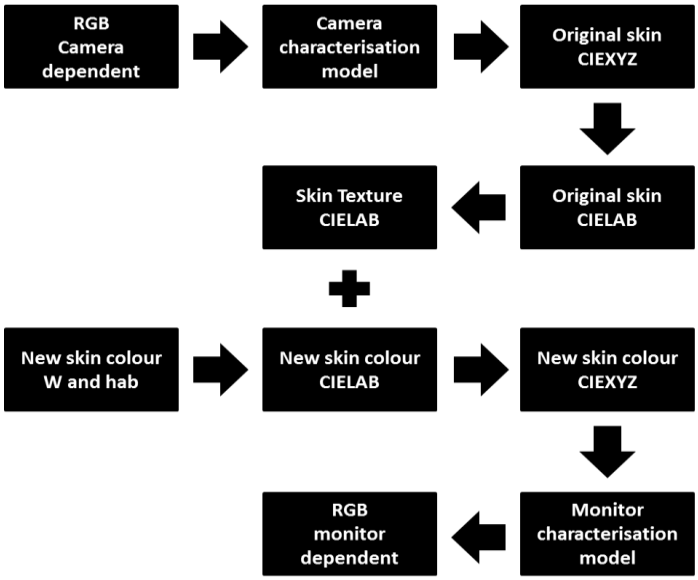


Figure B1 The data flow and structure of the code.

The skin colour of the original images

To achieve the skin whiteness trend line of the four subjects (a Chinese female, a Chinese male, a Caucasian female and a Caucasian male), the L^* and C^*_{ab} of these four subjects' skin colour and the skin whiteness indices of the ethnic group they were in were needed (see Section 4.1.2.6). The L^* and C^*_{ab} of these four subjects (named $\boldsymbol{L^*_{sub}}$ and $\boldsymbol{C^*_{ab,sub}}$)were listed in the table below (Table B2). These values were calculated from the average CIELAB of the four facial locations that measured by using the PR650.

Table B2 The skin colour information of four original images.

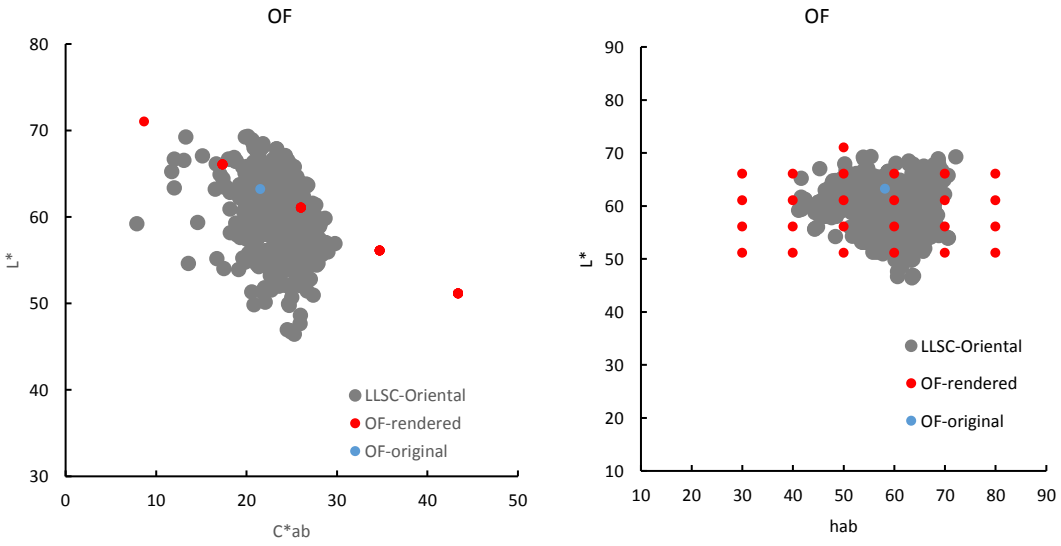
	$\boldsymbol{L^*_{sub}}$	$\boldsymbol{C^*_{ab,sub}}$
Caucasian Male original image	63.89	20.98
Caucasian Female original image	66.83	17.99
Chinese Male original image	57.18	24.85
Chinese Female original image	63.74	21.38

The formulae of the skin trend line of the four subjects can be achieved, as listed in the table below (Table 8).

Table B3 The skin whiteness trend line of the four subjects in the stimuli images.

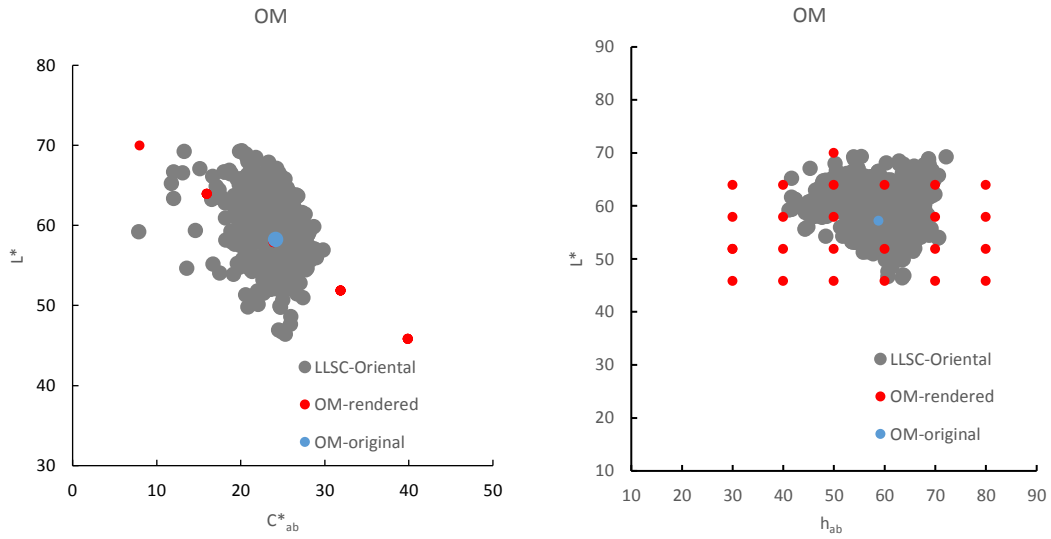
	Skin Whiteness trend line function
Caucasian Male	y = -0.8009x + 80.7
Caucasian Female	y = -0.7706x + 80.7
Chinese Male	y = -0.7571x + 76
Chinese Female	y = -0.5732x + 76

With these skin whiteness trend line functions, the selected skin whiteness and skin hue angle can be transformed into CIELAB. Figure 2 –Figure 5 shows the distribution of the selected skin colours for examination in the CIELAB $L^*C^*_{ab}$ and the CIELAB a^*b^* plane.



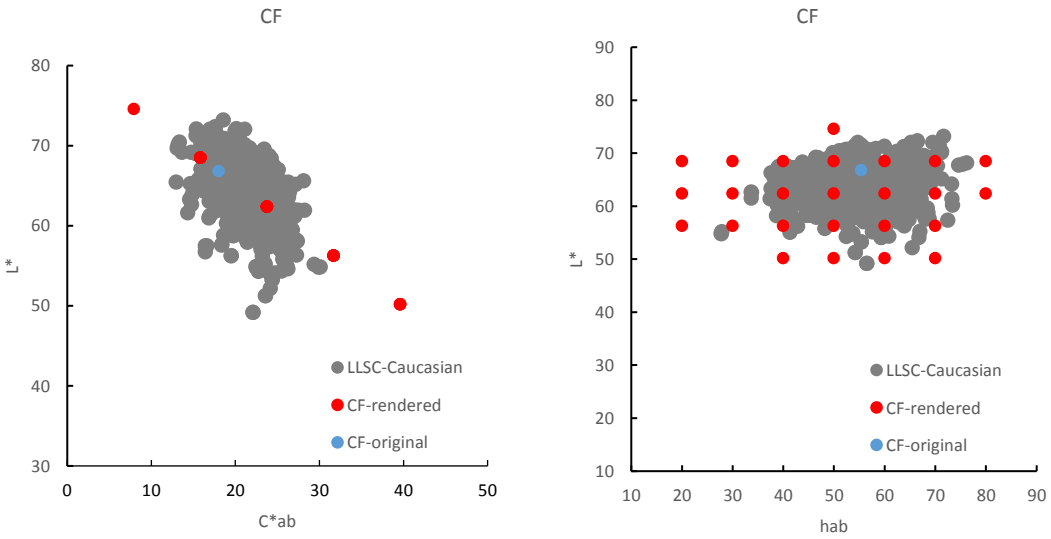
(a) (b)

Figure B2 The distribution of the selected skin colours of OF for examination: (a) in $\boldsymbol{L^*C^*_{ab}}$ plane; (b) in $\boldsymbol{L^*h_{ab}}$ plane.



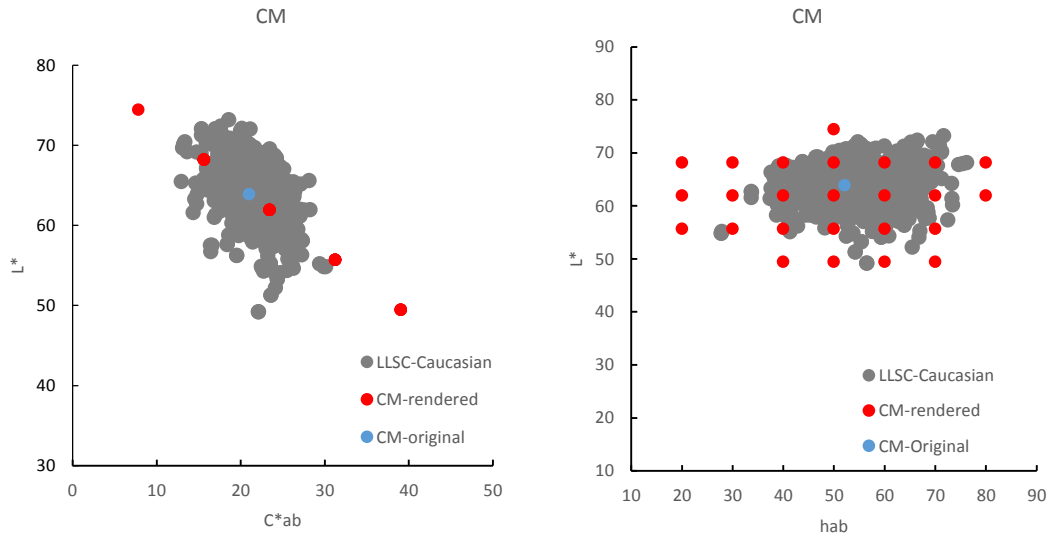
(a) (b)

Figure B3 The distribution of the selected skin colours of OM for examination: (a) in the $L^*C^*_{ab}$ plane; (b) in the L^*h_{ab} plane.



(a) (b)

Figure B4 The distribution of the selected skin colours of CF for examination: (a) in the $L^*C^*_{ab}$ plane; (b) in the L^*h_{ab} plane.



(a) (b)

Figure B5 The distribution of the selected skin colours of CM for examination: (a) in the $\boldsymbol{L^*C^*_{ab}}$ plane; (b) in the $\boldsymbol{L^*h_{ab}}$ plane.

Transform u'v' to CIEXYZ

In Yano, *et al*'s research, the preferred Japanese skin colour were reported in u'v'. To cross compare the skin colour preference between Yano and the present study, their results need to transform into CIEXYZ, then transform into CIELAB. Here, the lightness of Yano's results was assumed to equal to 60. According to Equation 2.5, the Y coordinate of the CIEXYZ can be achieved via the Equation B1.

$$\mathrm{Y} = Y_n \times \left(\frac{L^* + 16}{116}\right)^3 \qquad \textbf{Equation B1}$$

where $L^* = 60$ and Y_n=100.

Then the Y coordinate can be calculated, which equal to 28.12.

The u' and v' coordinates can be transformed into x and y coordinated via solving the two equations in Equation 2.4. Equation B2 shows the solution.

$$x = \frac{9u'}{6u' - 16v' + 12}$$

$$y = \frac{4v'}{6u' - 16v' + 12}$$

Equation B2

Then the x and y can be gained which equal to 0.3881 and 0.3482.

The Y, x and y coordinates can be transformed into CIEXYZ by solving the Equation 2.3. The solutions are listed in Equation B3

$$\mathrm{X} = x\frac{Y}{y}$$

$$\mathrm{Y} = Y$$

$$\mathrm{Z} = (1 - x - y)\frac{Y}{y}$$

Equation B3

Then, the X, Y and Z can be achieved, which were 31.35, 28.13 and 21.30. The CIEXYZ then can transform into CIELAB via Equation 2.5.

Peak cancellation

Figure B6 showed the SPD of the light source (marked in blue) and SPD of the skin colour a subject (marked in red). It can be seen from this figure that both SPD have the peaks at about the 435 nm and 545 nm.

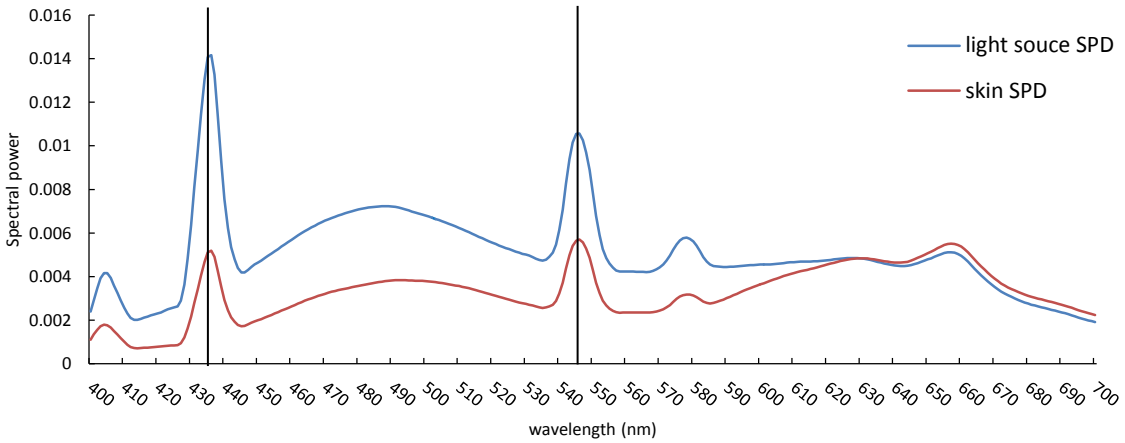


Figure B6 The SPD of the light source and the the SPD of the skin

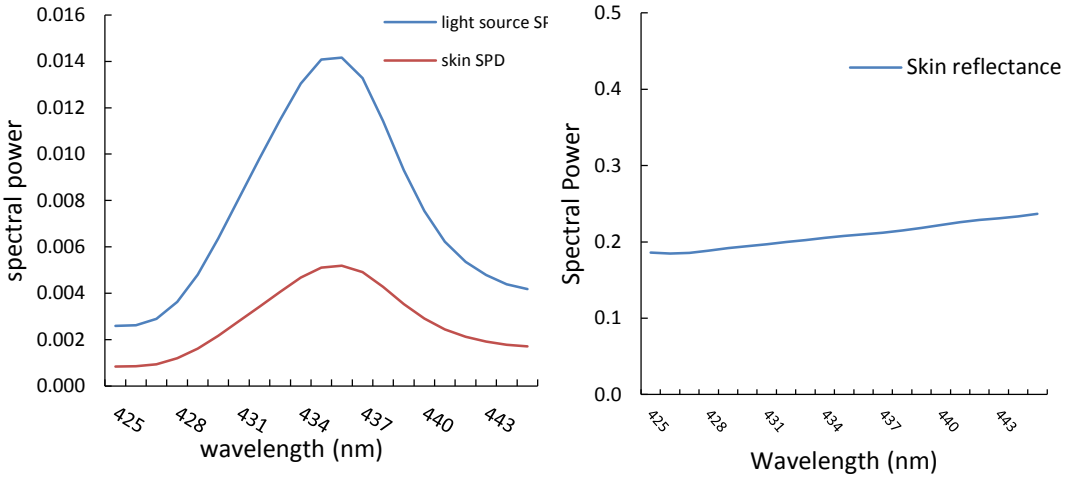
According to Equation 3.1, the reflectance of the skin colour can be calculated by using the SPD of the skin divide the SPD of the light source. With the mathematical correction of the reference white, the Equation 3.1 can be rewritten as shown below:

$$R_{\text{sample}} = R_w \times \frac{SPD_{sample}}{SPD_w} \qquad \textbf{Equation B4}$$

According to the equation, the Table B4 showed the fraction of the calculation process of SPD_{sample}/SPD_w. The section of the data from 425nm to 445nm were selected, as one of the peak is included in this wavelength range. The calculated results were also plotted in the Figure B7 (b).

Table B4 The data of the wavelength range of 425nm to 445nm

WAVELENGTH (NM)	LIGHT SOURCE SPD	SKIN SPD	SKIN REFLECTANCE
425	2.59E-03	8.46E-04	1.86E-01
426	2.63E-03	8.52E-04	1.85E-01
427	2.90E-03	9.43E-04	1.86E-01
428	3.63E-03	1.20E-03	1.89E-01
429	4.80E-03	1.61E-03	1.92E-01
430	6.36E-03	2.17E-03	1.94E-01
431	8.11E-03	2.80E-03	1.97E-01
432	9.81E-03	3.43E-03	2.00E-01
433	1.15E-02	4.06E-03	2.02E-01
434	1.30E-02	4.68E-03	2.05E-01
435	1.41E-02	5.11E-03	2.08E-01
436	1.42E-02	5.19E-03	2.10E-01
437	1.33E-02	4.91E-03	2.12E-01
438	1.14E-02	4.28E-03	2.15E-01
439	9.30E-03	3.54E-03	2.18E-01
440	7.55E-03	2.92E-03	2.22E-01
441	6.23E-03	2.45E-03	2.26E-01
442	5.35E-03	2.13E-03	2.29E-01
443	4.79E-03	1.92E-03	2.31E-01
444	4.39E-03	1.78E-03	2.34E-01
445	4.18E-03	1.72E-03	2.37E-01



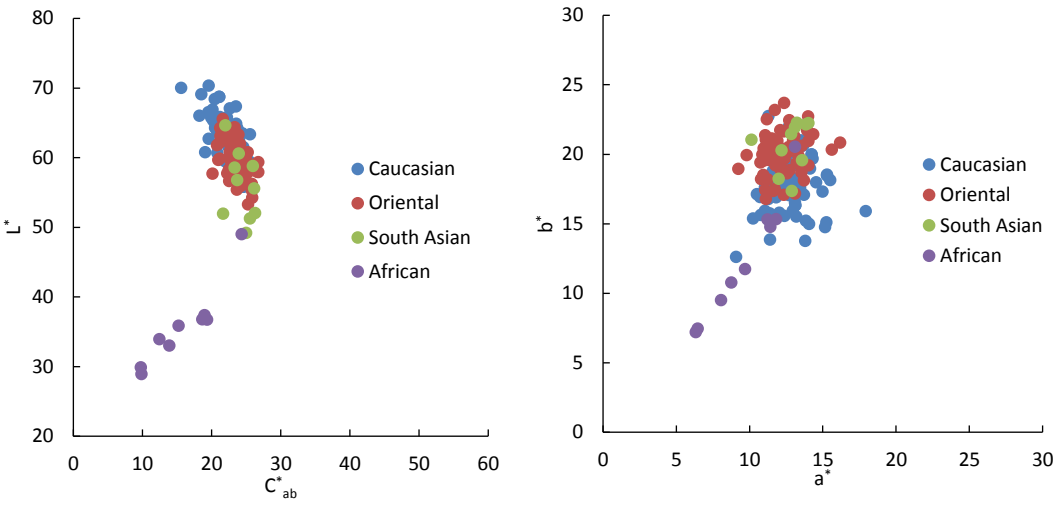
(a) (b)

Figure B7 The example of the peak in 435nm.(a) the section of the SPD of the light source and skin; (b) the reflectance of the skin calculated from the SPD in (a).

Comparing the Figure B7 (a) and (b) can find that the peak is cancelled after the calculation.

The plots of the facial skin colour (PR650)

the facial skin colour of all 188 subjects were plotted in a^*b^* and $L^*C^*_{ab}$ planes, as shown in Figure B8, to investigate the variation of the subjects.



(a) (b)

Figure B8 The plot of the average facial skin colour.(a) in $L^*C^*_{ab}$ plane;(b) in a^*b^* plane

Appendix C

The comparison of two experimental designs: 5 question in one trial and one question in one trial

To investigate the agreement between two experimental designs, 20 pairs of images (included 10 pairs of Oriental Female and 10 pairs of Oriental males) were used. Ten Chinese observers were volunteered to participate this test. The observers were asked to do two tests on two different days with one week (5 days) time gap, which ensures the observer to had limited memory of the previous test. The Cohen's kappa is a statistical method that frequently used to determine the significance of the difference between different experimental designs (Burns and Dobson, 2012). Here, the Cohen's kappa was used to determine the difference between answer one question each trial and answer five question in one triall, as shown in Figure C1 and Figure C2, respectively. The SPSS was used to calculate the Kappa (Howitt and Cramer, 2011).

Figure C1 The interface of the experimental design of one question each triall.

Figure C2 The interface of the experimental design of five questions each trial.

Seven Chinese university students participated this test. The test results were listed in Table C1.

Table C1 The Cohen's kappa value of seven participants.

Observer number	Cohen's kappa
Observer 1	0.448
Observer 2	0.552
Observer 3	0.582
Observer 4	0.803
Observer 5	0.843
Observer 6	0.694
Observer 7	0.524
mean	0.592

It can be known from Table 9 that the average Cohen's kappa value is 0.592. The Cohen's kappa value from 0.40 to 0.60 is considered as has a moderate agreement (Viera and Garrett, 2005; Hallgren, 2012).

Table C2 Rotated Component Matrix of the attributes in Component 2.

	Component
	1b
Rotation Sums of Squared Loadings % of Variance	78.459
Translucent	0.959
Fair	0.954
Smooth	0.945
Moist	0.930
Youth	0.910
Glossy	0.896
Clear	0.858
Intelligent	0.793
Pink	-0.690

Appendix D

Agreement between female and male Chinese observers in line chart

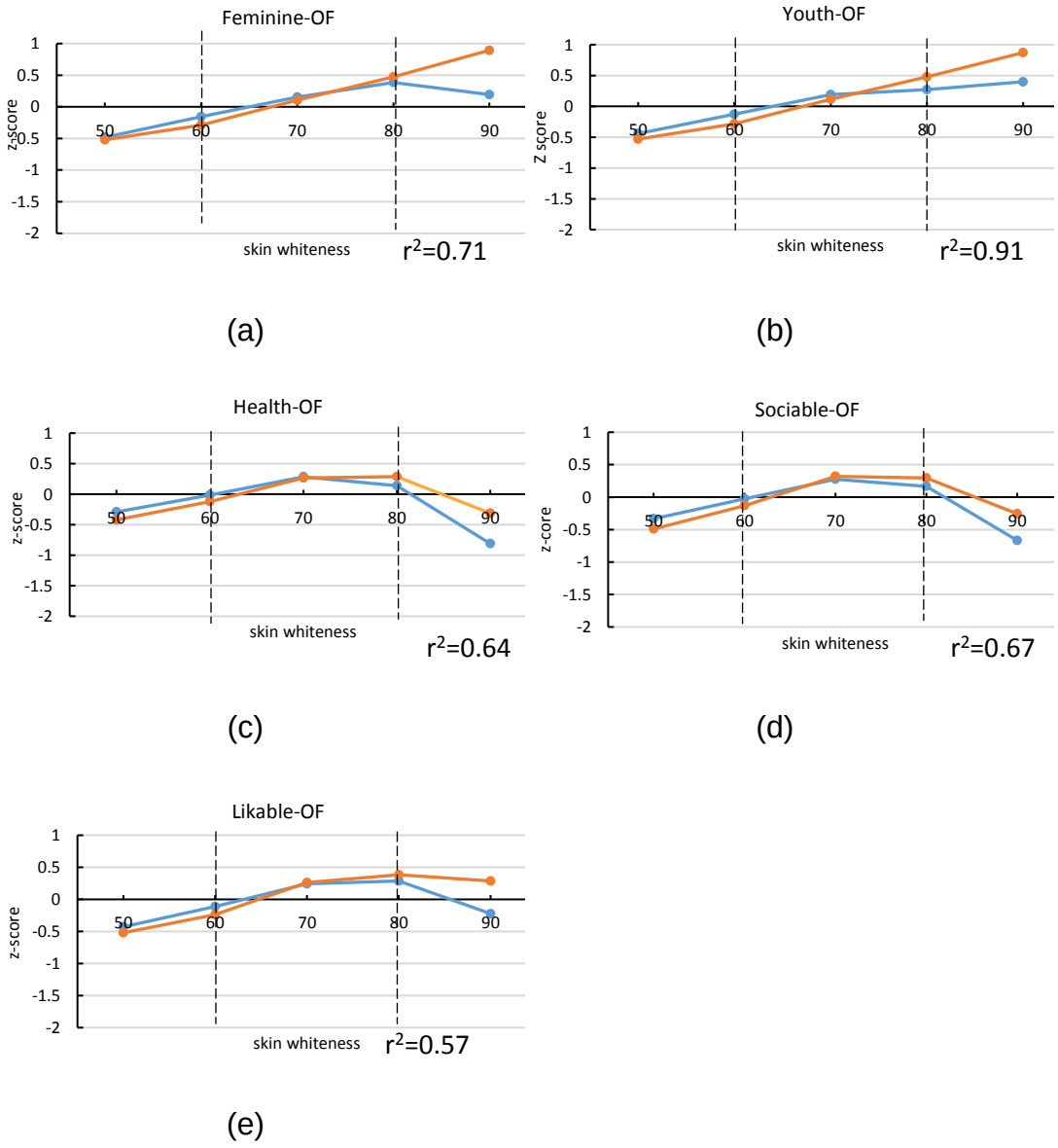


Figure D1 The judgements of the OF images with different whiteness from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

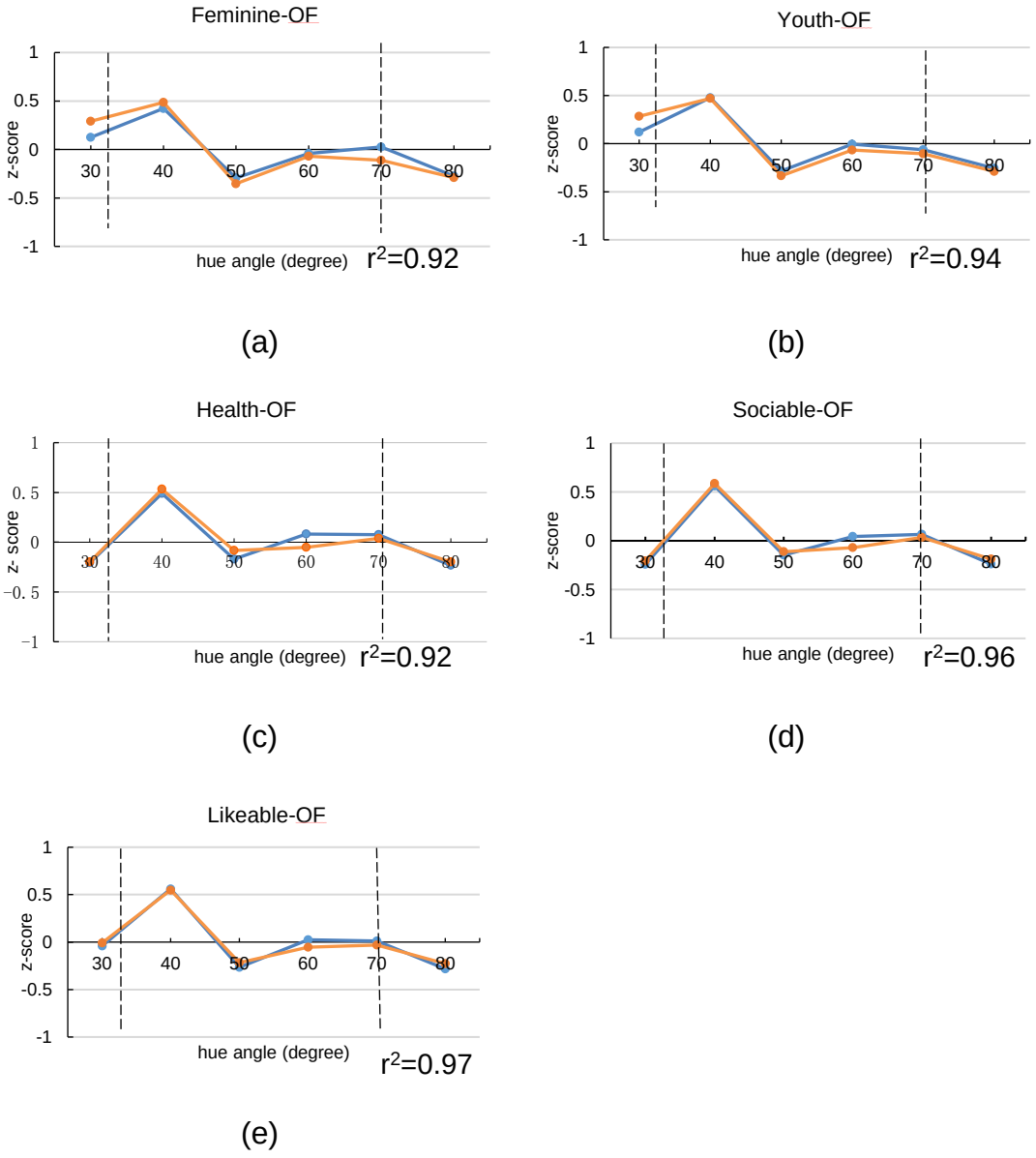


Figure D2 The judgements of the OF images with different hue angles from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

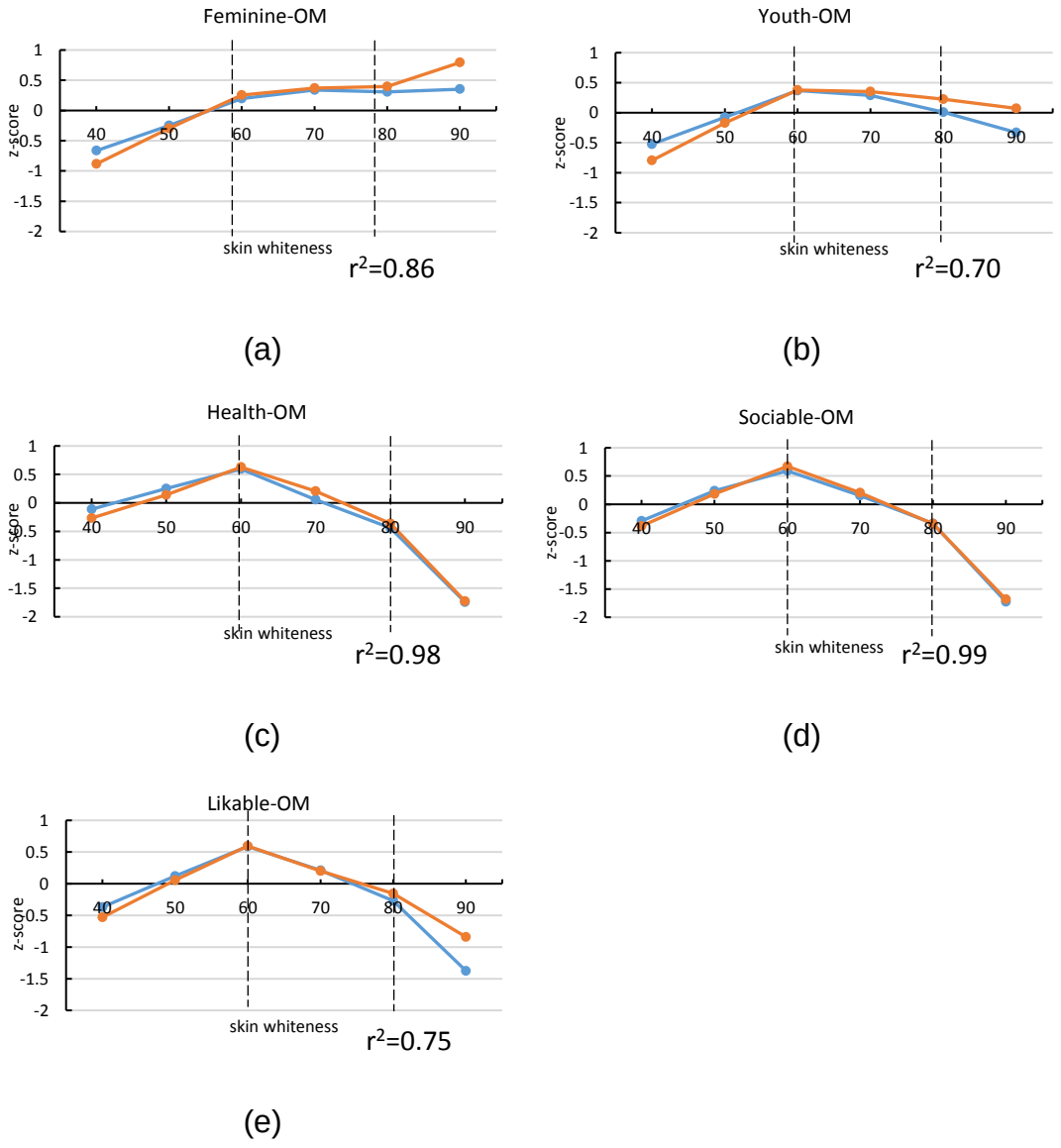


Figure D3 The judgements of the OM images with different whiteness from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

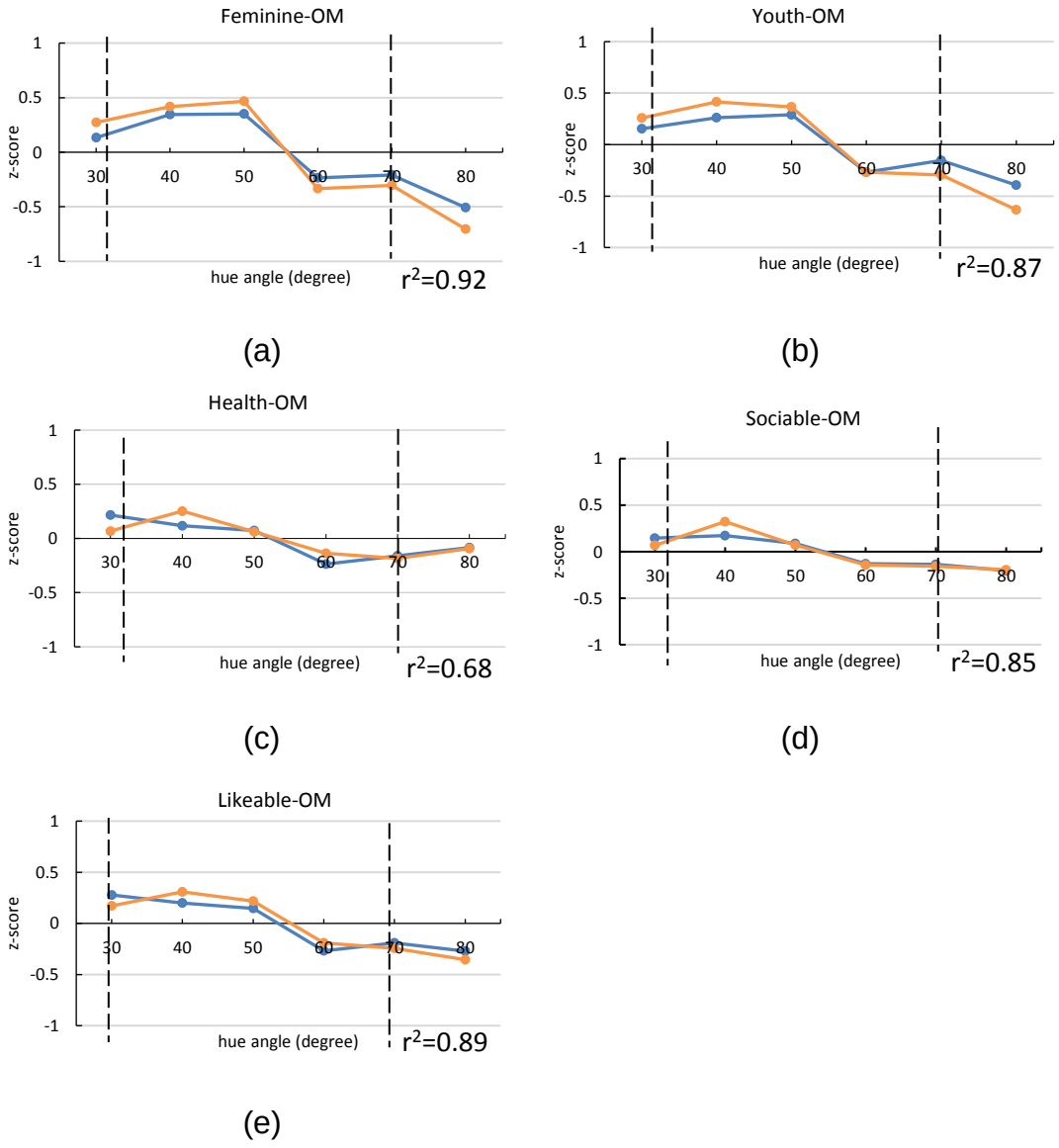


Figure D4 The judgements of the OM images with different hue angle from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

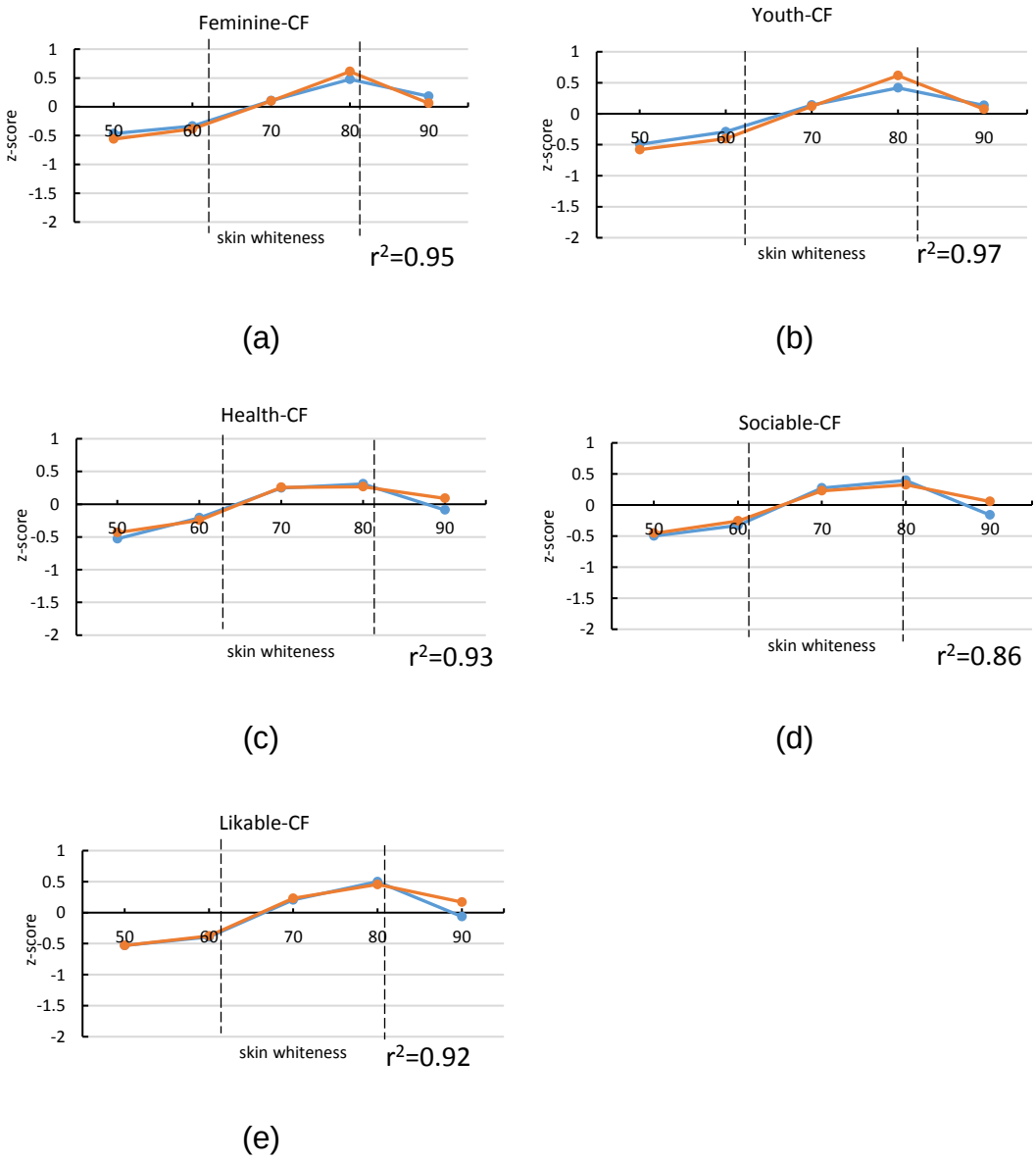


Figure D5 The judgements of the CF images with different skin whiteness from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

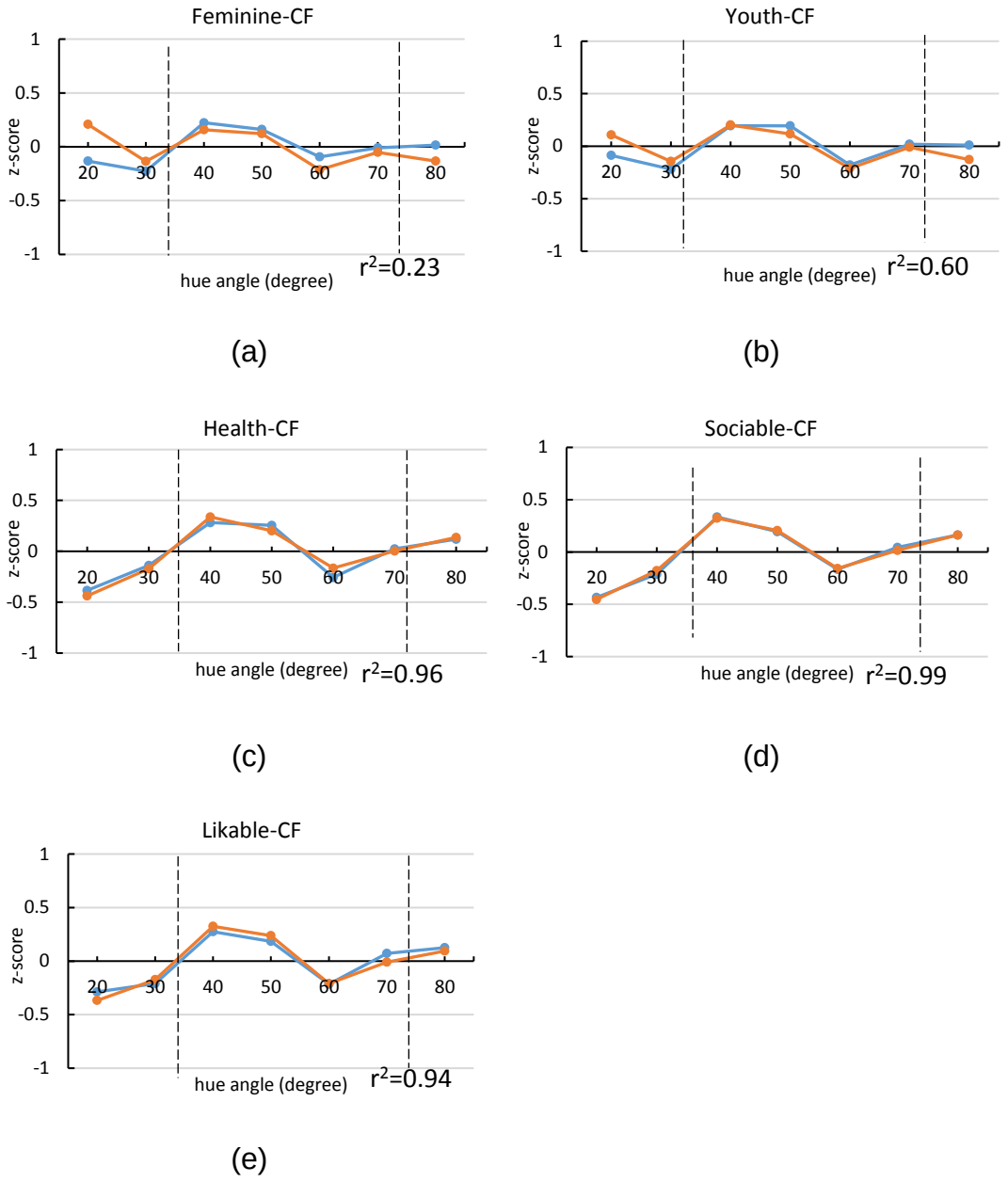


Figure D6 The judgements of the CF images with different hue angle from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

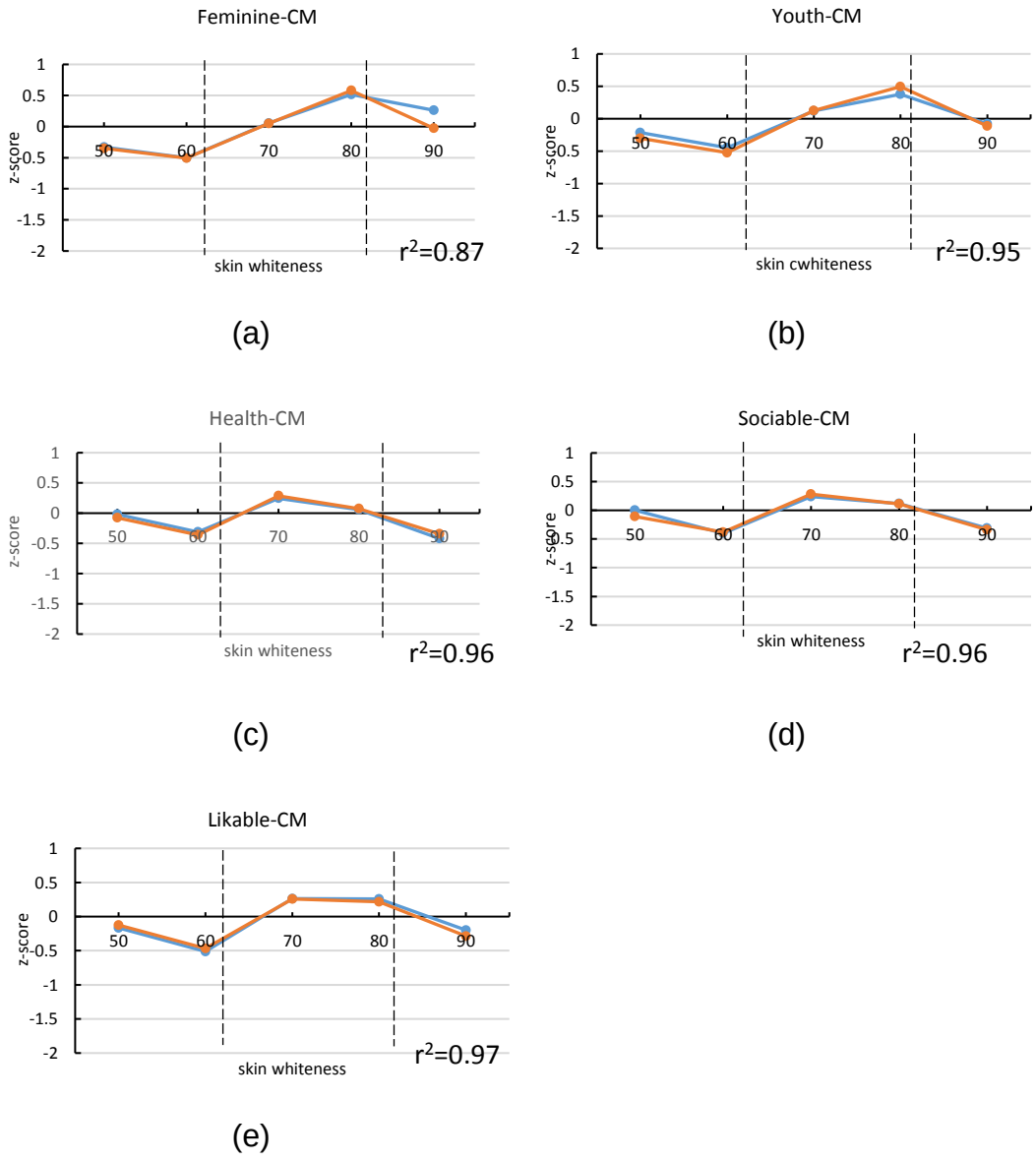


Figure D7 The judgements of the CM images with different skin whiteness from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

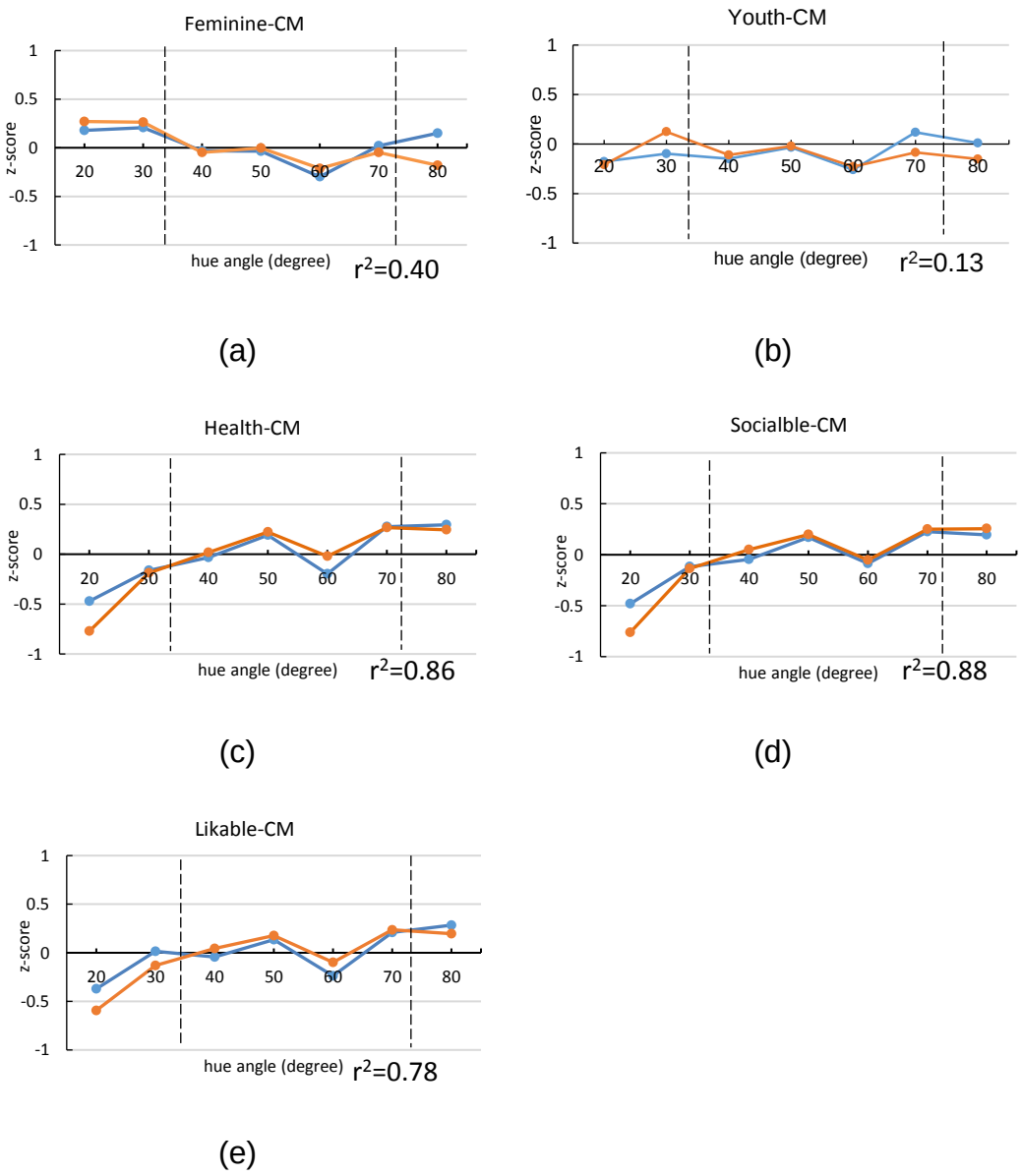


Figure D8 The judgements of the CM images with different hue angle from the male observers (blue line) and female observers (orange line) at five attributes: (a) Feminine; (b) Youth; (c) Health; (d) Sociable; (e) Likeable.

Agreement between the female and male observers in scatter diagram

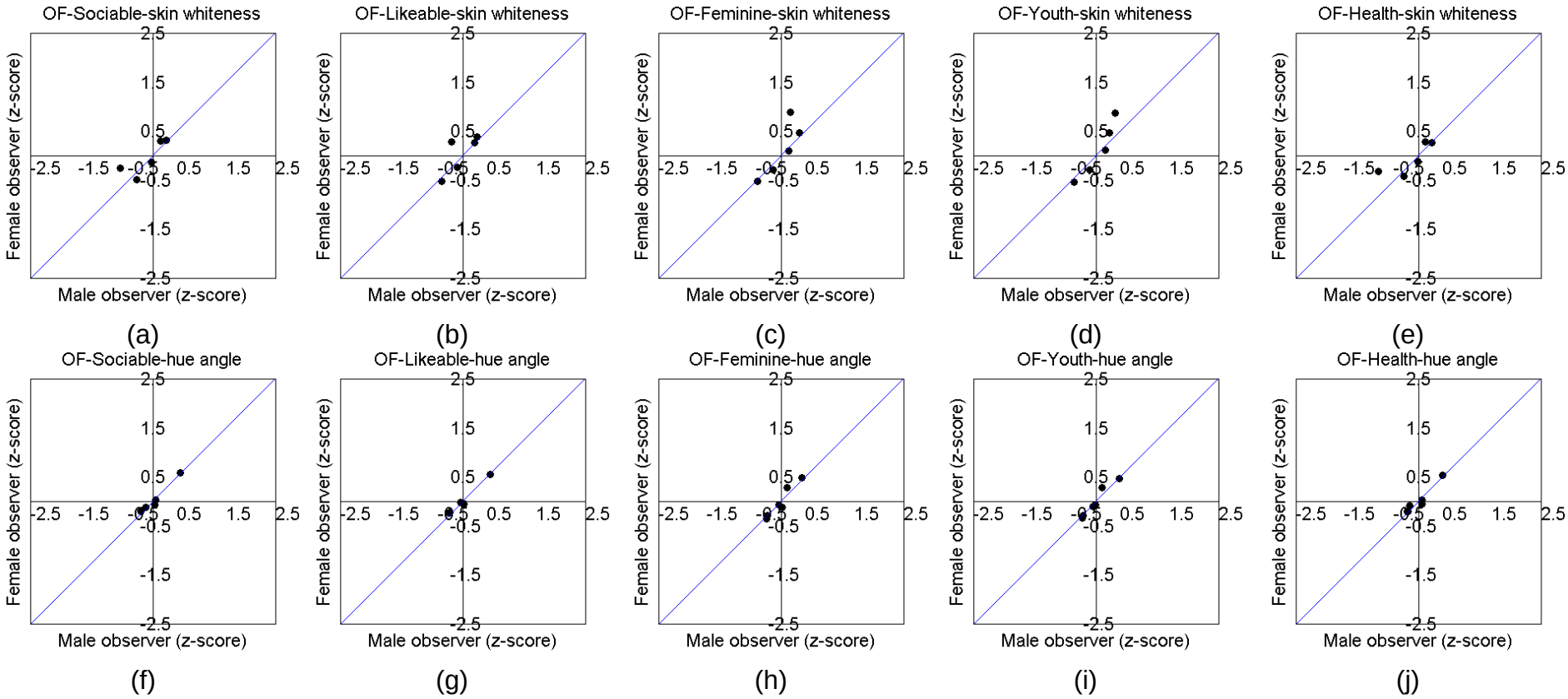


Figure D9 The agreement of the visual results of female observers against that of the male observers when observing OF image with different skin whiteness: (a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. and with different hue angle: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.

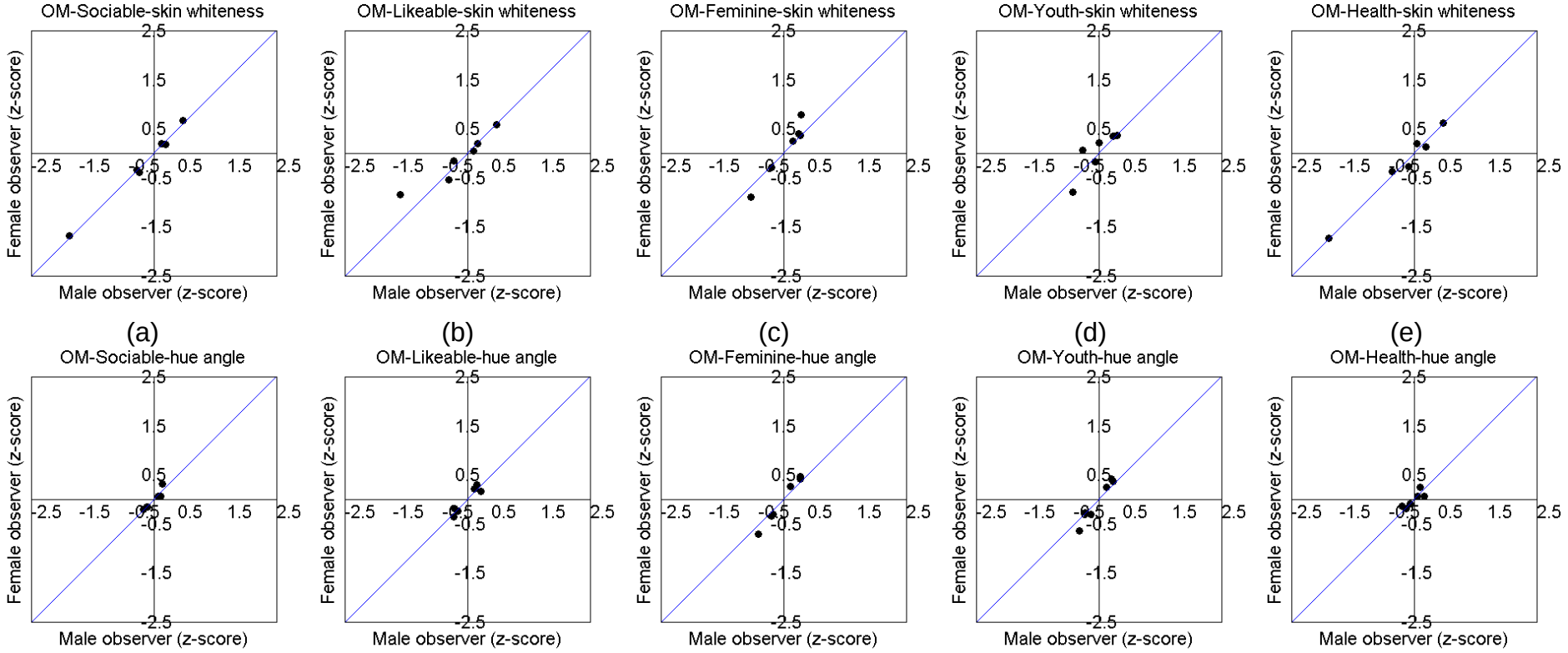


Figure D10 The agreement of the visual results of female observers against that of the male observers when observing OM image with different skin whiteness: (a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. and with different hue angle: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.

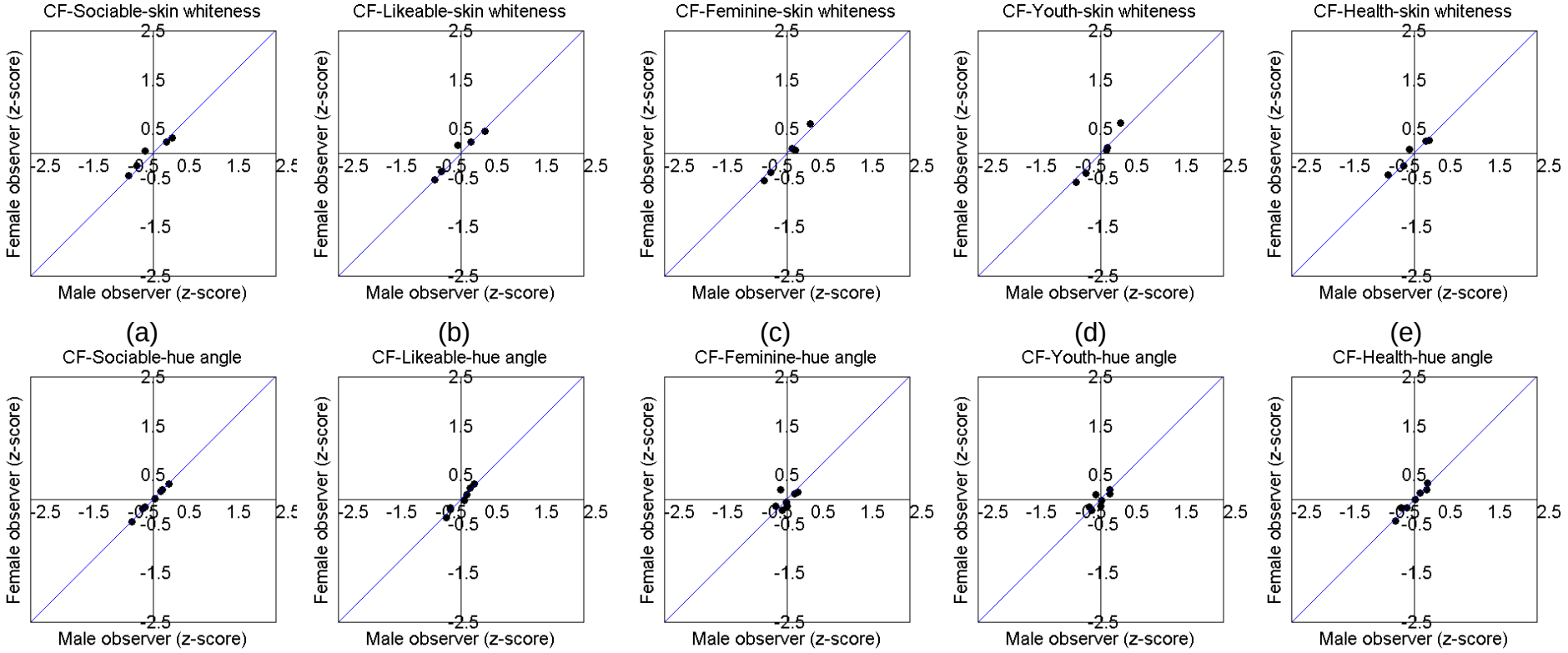


Figure D11 The agreement of the visual results of female observers against that of the male observers when observing CF image with different skin whiteness: (a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. and with different hue angle: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.

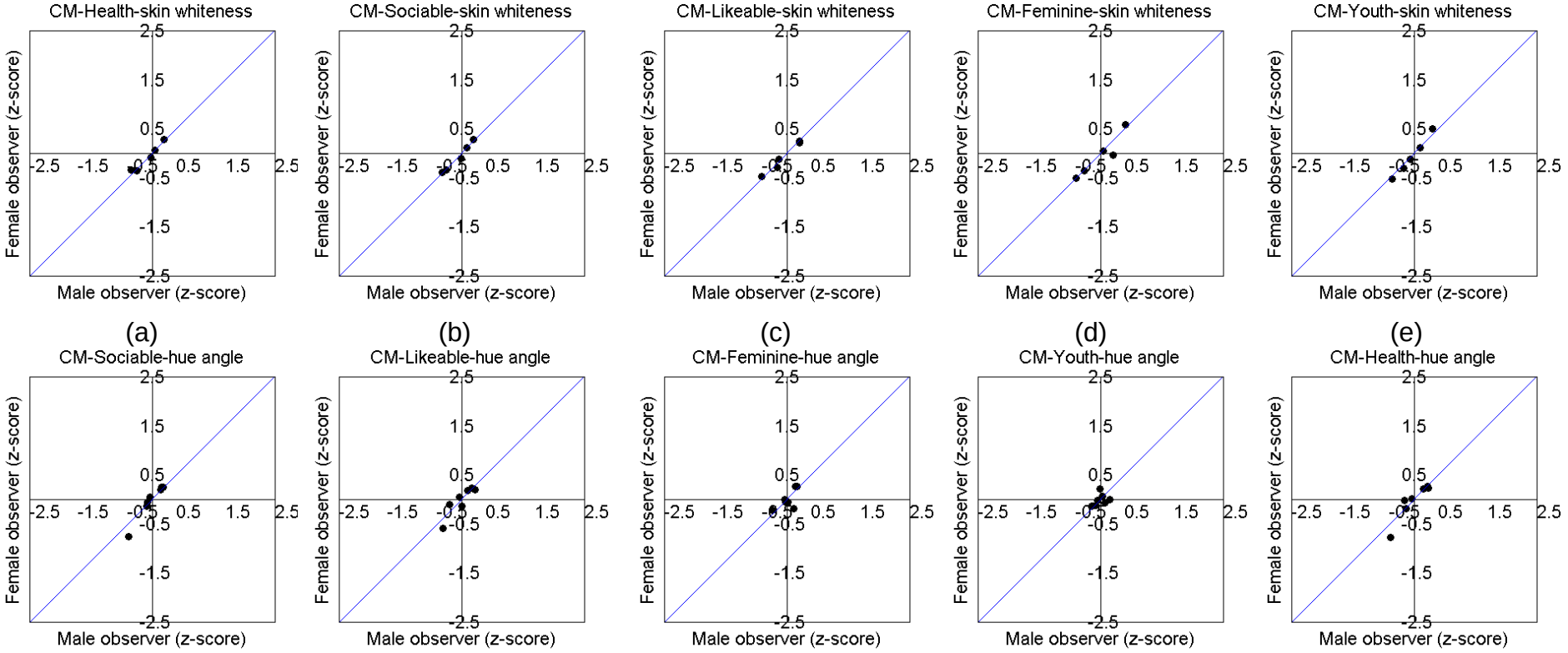


Figure D12 The agreement of the visual results of female observers against that of the male observers when observing CM image with different skin whiteness: (a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. and with different hue angle: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.

The correlations between every two impression attributes

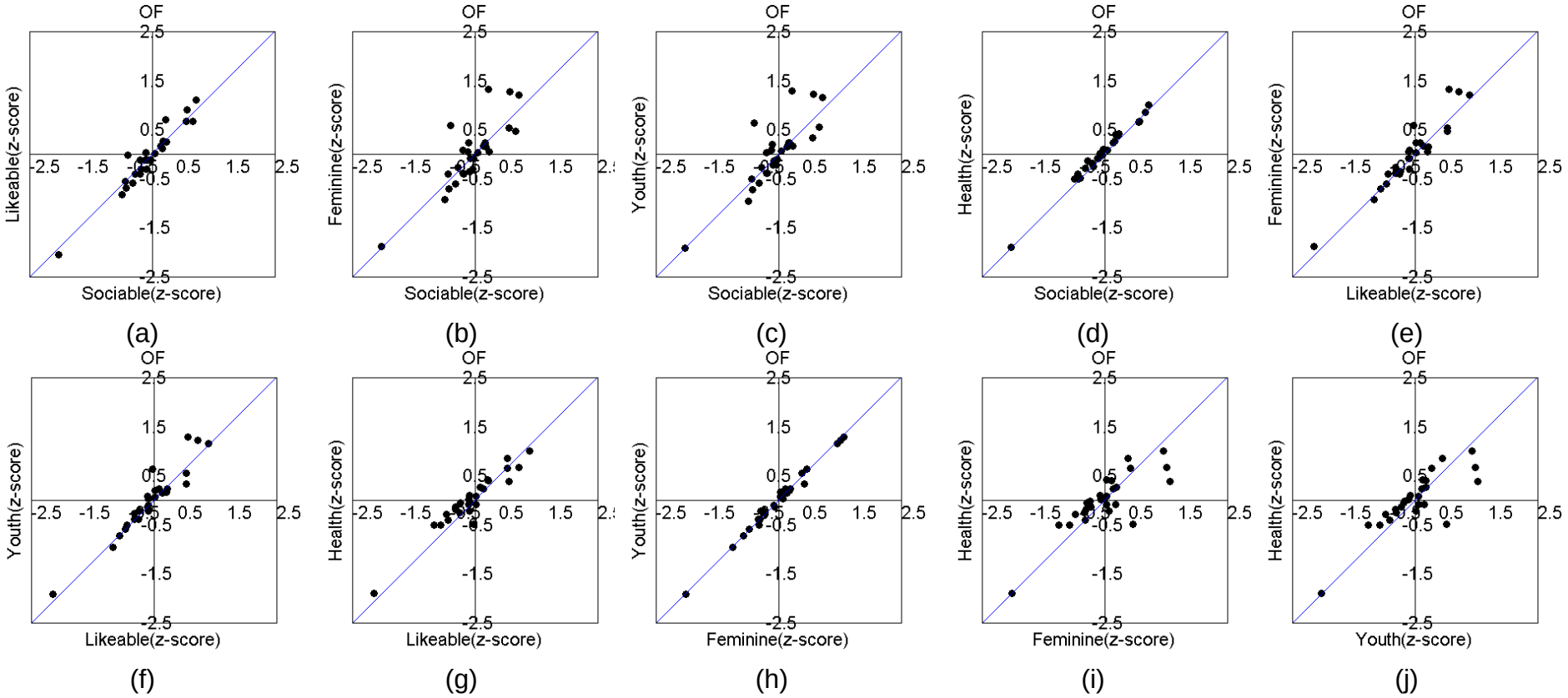


Figure D13 The agreement between two impression attributes at observing OF images (a) Sociable-likeable; (b) Likeable-Sociable; (c) Youth-Sociable; (d) Health-Sociable; (e) Feminine-Likeable; (f) Youth-Likeable (g) Health-Likeable; (h) Youth-Feminine; (i) Health-Youth.

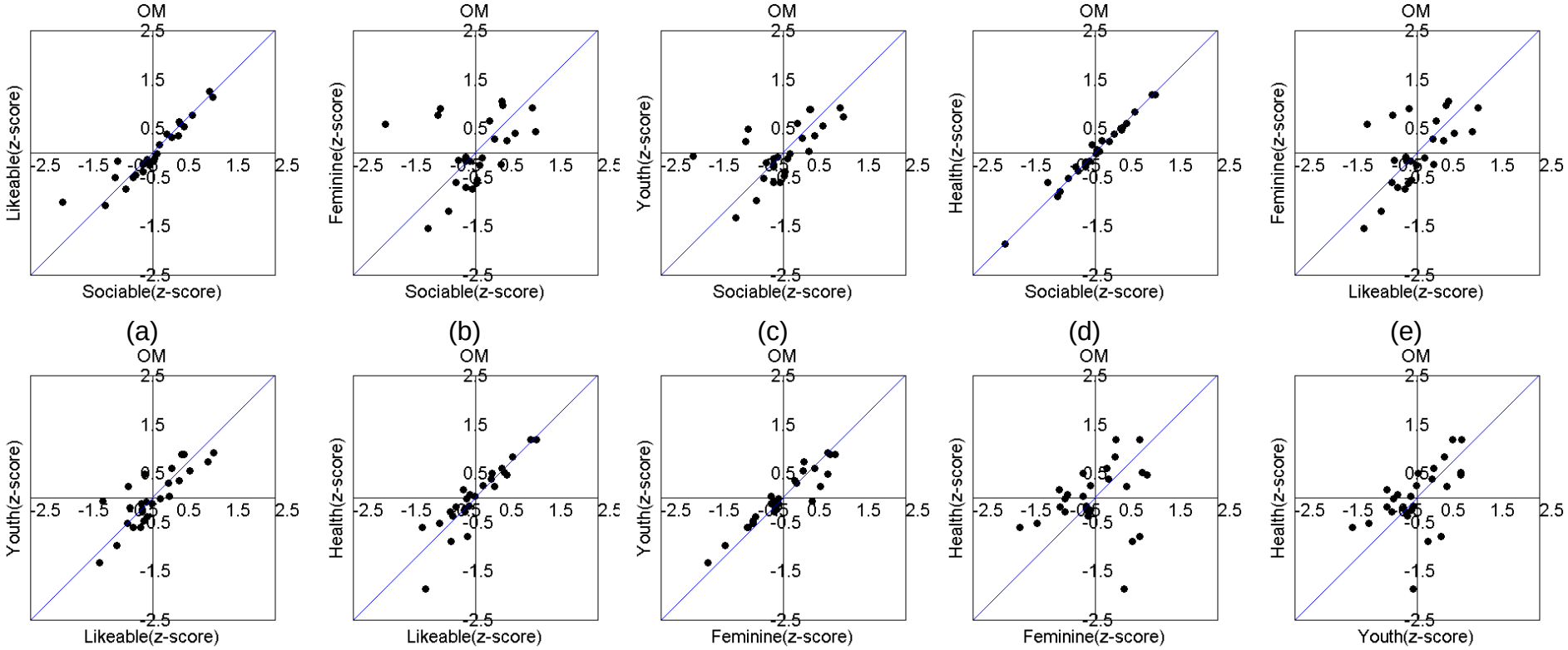


Figure D14 The agreement between two impression attributes at observing OM images (a) Sociable-likeable; (b) Likeable-Sociable; (c) Youth-Sociable; (d) Health-Sociable; (e) Feminine-Likeable; (f) Youth-Likeable (g) Health-Likeable; (h) Youth-Feminine; (i) Health-Youth.

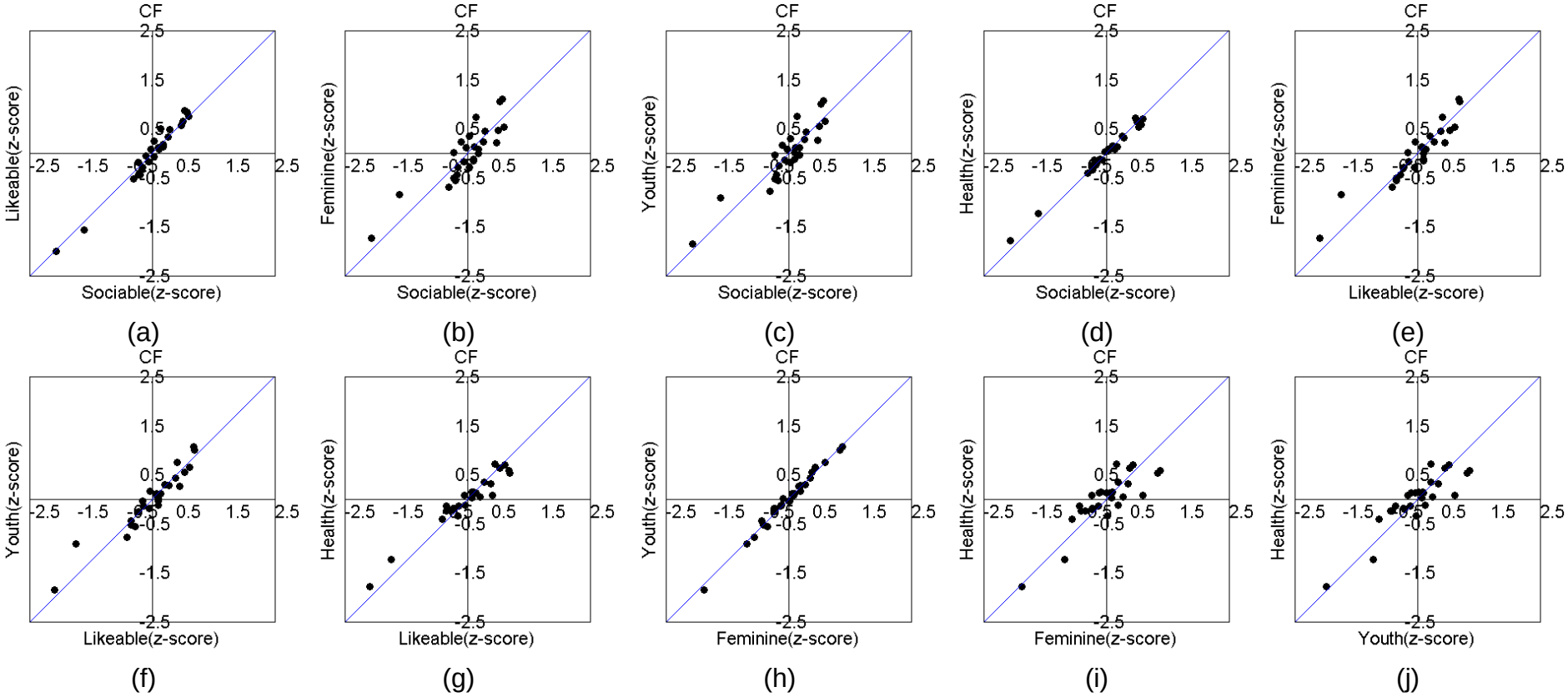


Figure D15 The agreement between two impression attributes at observing CF images (a) Sociable-likeable; (b) Likeable-Sociable; (c) Youth-Sociable; (d) Health-Sociable; (e) Feminine-Likeable; (f) Youth-Likeable (g) Health-Likeable; (h) Youth-Feminine; (i) Health-Youth.

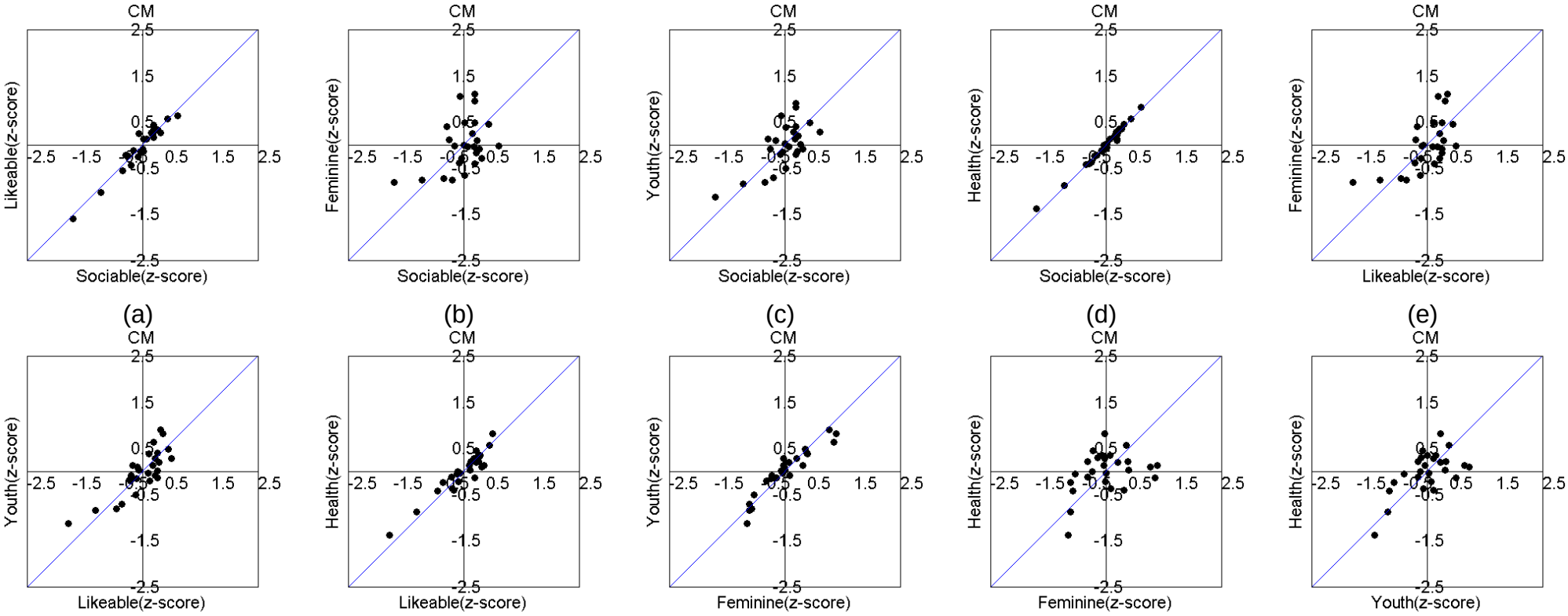


Figure D16 The agreement between two impression attributes at observing CM images (a) Sociable-likeable; (b) Likeable-Sociable; (c) Youth-Sociable; (d) Health-Sociable; (e) Feminine-Likeable; (f) Youth-Likeable (g) Health-Likeable; (h) Youth-Feminine; (i) Health-Youth.

Test results of the developed models

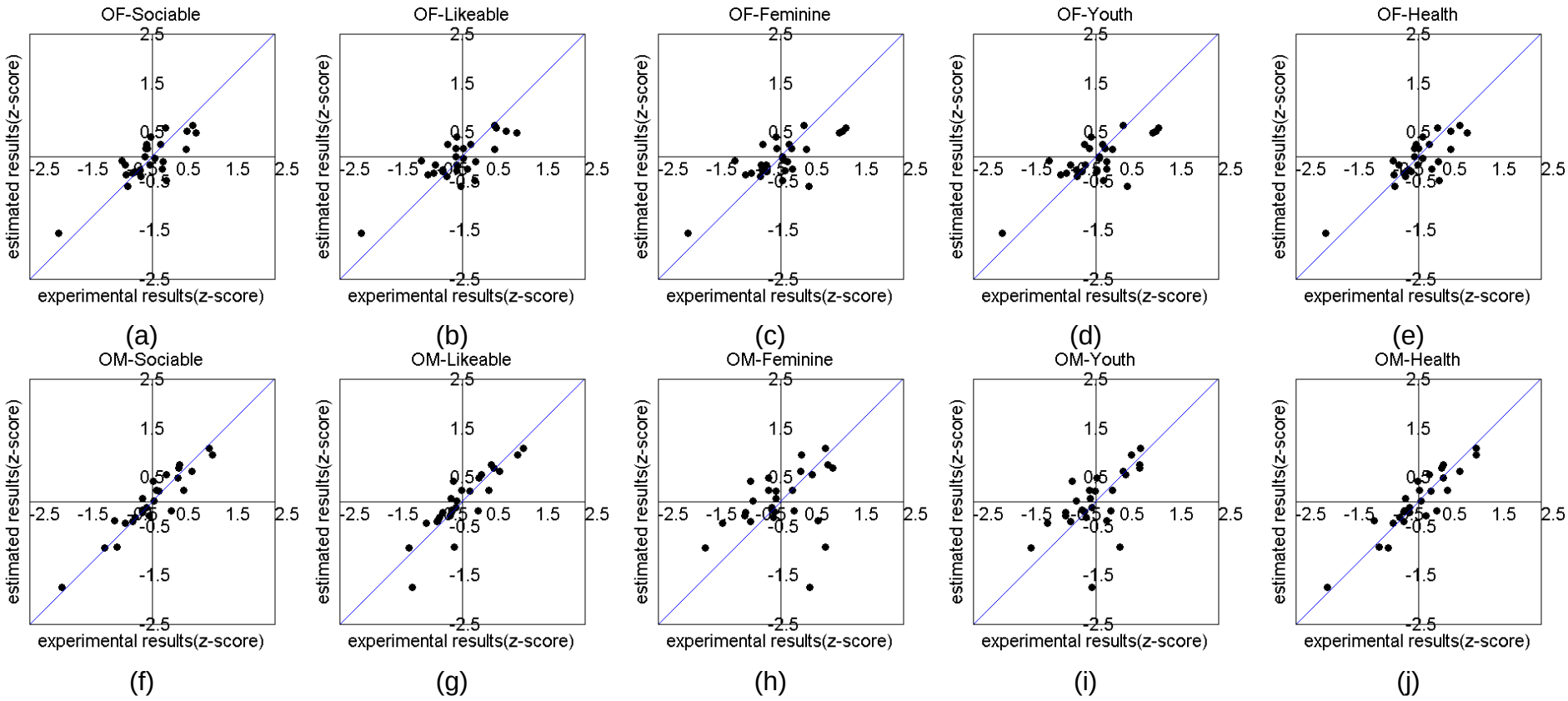


Figure D16 The agreement between the visual results and the modelling results when observing OF images :(a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. And when observing OM images: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.

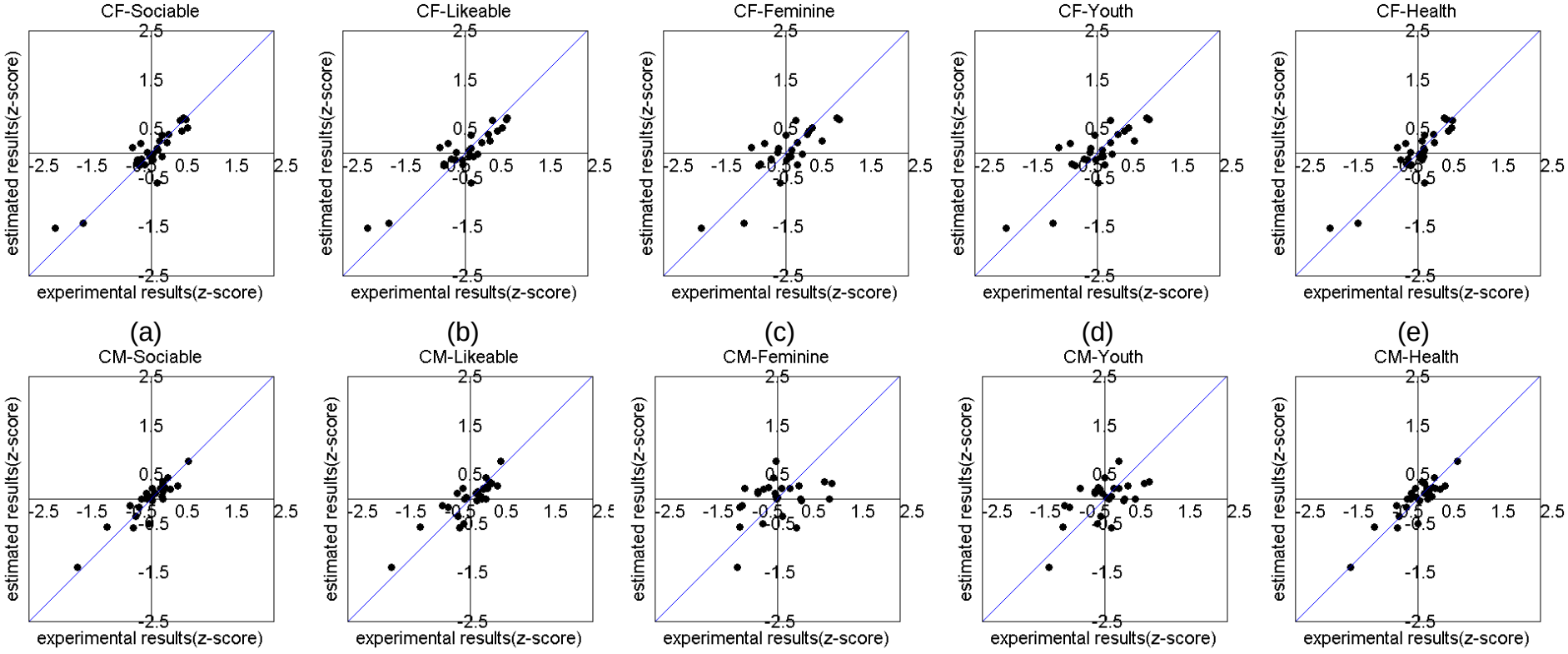


Figure D17 The agreement between the visual results and the modelling results when observing CF images :(a) Sociable; (b) Likeable; (c) Feminine; (d)Youth; (e) Health. And when observing CM images: (f) Sociable; (g) Likeable; (h) Feminine; (i)Youth; (j) Health.