Model Facial Colour Appearance and Facial Attractiveness for Human Complexions

Yan Lu

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Design

March 2023

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 publications related to Chapter 3 and Chapter 4:

Lu, Y., Yang, J., Xiao, K., Pointer, M., Li, C. and Wuerger, S. 2021. Skin coloration is a culturally-specific cue for attractiveness, healthiness, and youthfulness in observers of Chinese and western European descent A. Jones, ed. PLOS ONE. 16(10), p.e0259276.

Lu, Y., Yang, J., Xiao, K., Pointer, M., Li, C. and Wuerger, S. 2020. Investigation of effect of skin tone to facial attractiveness. 28th Color and Imaging Conference Final Program and Proceedings. Proceedings of IS&T 28th Color and Imaging Conference Society for Imaging Science and Technology. 2020(28), pp.25–29.

Lu Y; Yang J; Xiao K; Pointer M; Li C; Wuerger S. 2021 Skin Colour is a Culturally-Specific Cue for Attractiveness, Healthiness and Youthfulness in Chinese and Caucasian Observers. AVA 2021 Spring Meeting

The publications related to Chapter 3 and Chapter 5:

Lu, Y., Xiao, K., Yang, J., Pointer, M., Li, C. and Wuerger, S., 2022. Different colour predictions of facial preference by Caucasian and Chinese observers. Scientific Reports, 12(1), p.12194.

Lu, Y., Xiao, K., Yang, J., Pointer, M., Li, C. and Wuerger, S. 2022. Colour predictors of facial preference differ in Caucasian and Chinese populations. Journal of Vision. 22(14), p.3121.

The publications related to Chapter 3 and Chapter 7:

Lu, Y. and Xiao, K. 2021. Quantifying facial colour appearance of Caucasian and Chinese faces In: Proceedings of the AIC 14th Congress 2021 International Colour Association (AIC)., pp.1233–1238.

Acknowledgements

Abstract

Human facial complexion has been a subject of great interest in many areas of science and technology including dermatology, cosmetology, computer graphics, and computer vision. Facial colour appearance conveys vital personal information and influences social interactions and mate choices as contributing factors to perceived beauty, health, and age. How various colour characteristics affect facial preference and whether there are cultural differences are not fully understood. On the other hand, facial colour appearance cannot be simply quantified by colour measurement. Facial colour perception is distinctive. The perceptual aspects of facial colour appearance haven't been precisely investigated.

The present study aims to better understand the human colour perception of facial complexions. Psychophysical experiments were carried out to assess facial colour preference and facial colour appearance, respectively. A set of facial images of real human faces were used and the colour was rigorously controlled in those experiments so that the facial colour appearance could be evaluated based on the realistic skin models.

Experiments on colour preference provided a thorough assessment of the relationships between various facial colour characteristics and preference judgements and meanwhile revealed large cultural differences between Caucasian and Chinese populations. A useful and repeatable analytical framework for facial preference modelling was provided. This work contributes to the growing body of research using realistic skin models and highlights the importance of examining various colour cues utilized in facial preference evaluation.

Experiments on colour appearance for the first time precisely measured the overall colour perception of facial appearance. New indices WIS, RIS, and YIS were developed to accurately quantify perceived facial whiteness, redness, and yellowness. The perceptual difference between the colour appearance of the face stimuli and the nonface stimuli was discovered.

Taken together, the present study shed new light on how our visual system perceives and processes colour information on human faces.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Colour is the most common but crucial visual sensation to humans. As one of the most significant features of human faces, skin colour conveys vital personal information and influences the impression and preference perceived by others. The relationship between skin colour and facial preference is of great importance in numerous applications and occasions where effort has been made to satisfy people's desire to have a beautiful, healthy-looking or youthful facial appearance. For example, facial colour preference may influence the product expectations in the cosmetic industry, the aesthetic facial criteria in plastic surgeries, preference-based facial colour reproduction in the imaging industry, and the colour rendition properties in the lighting industry, etc (Zeng and Luo, 2010; Xie and Zhang, 2013; Gao et al., 2018).

Studies on facial impressions and preference judgements started early from the structural facial traits such as facial symmetry, averageness and sexual dimorphism (Thornhill and Gangestad, 1999; Rhodes, 2006). Compared to those non-colour-related facial traits, the colour appearance of human faces has been relatively less investigated but has gained increasing attention in the last ten to twenty years, which may suggest an important role for facial colour information in any of the preference-related judgements including facial attractiveness, perceived healthiness, and perceived age. Colour appearance is also easier to change compared to those structural determinants. It can change either slowly and continuously due to UV exposure (Amano et al., 2020), fruit and vegetable (FV) consumption (Tan et al., 2015) or rapidly and momentarily due to changes in the physical or emotional state (Bilal et al., 2015), use of coloured cosmetics, change in the lighting environments. Skin colour, and all these subtle colour variations, can be sensitively perceived by human observers (Changizi et al., 2006). Hence, understanding the psychological effect of colour on facial preference has profound implications under various social contexts and could benefit numerous applications (Elliot & Maier, 2014).

Several facial colour characteristics, such as overall facial redness, yellowness, colour variation, colour contrasts, etc. have been considered as individual crucial parameters of preference judgements. The axiom in most previous research is to change a single colour variable in a controlled experiment for preference evaluation, which neglects the holistic process of facial colour perception in real life. None of the studies has fully considered

different colour characteristics on human faces, and little is known about the impact of various colour cues taken together on facial preference judgment. More importantly, with those controlled experiments, facial preference has been studied based on colour-manipulated skin models rather than realistic skin models. It is difficult to assess the importance of colour in preference judgements on real human faces.

There are also large effects on cultural and environmental differences involved. Since people typically rely on their own perceptions to judge the facial impression or preferences, it is important to know the perceptual differences when diverse populations of both observers and the observed are involved. Different application fields will also need to focus on specific aspects based on the needs of people from different cultural backgrounds. Current studies have mostly been conducted among Caucasian populations, yet cross-cultural studies are limited and cultural differences between different ethnic groups are not satisfactorily understood.

On the other hand, the perception of facial colour appearance hasn't been precisely investigated. The facial colour appearance describes what the colour stimuli of facial skin look like in the human colour vision, which plays a key role in social perception processes (Thorstenson et al., 2020). There is an increasing number of applications that need to quantify skin colour appearance and reproduce it accurately. For example, accurate skin colour reproduction is considered as an indicator of good quality for many products in imaging industry such as the display and the camera (Imai et al., 1996); colour matching to the surrounding skin is extremely important in patients wearing maxillofacial prostheses (Xiao et al., 2013; Sohaib et al., 2018).

Studies on facial colour preference have commonly used average skin colour specified in CIELAB colour space to represent the overall facial colour appearance. Specifically, the mean L*, a* and b* values of the facial skin area are considered as the overall facial lightness, redness and yellowness, respectively. However, the relationship between these colour appearance attributes and the colorimetric values of a human face is not clear so far, and what is the overall colour appearance or the global colour impression of a human face remains unknown to us. Though the CIE colorimetry provides the objective tool for colour measurement and quantification of a certain point, our skin is a non-uniform multi-layered structure and such measurements cannot include the various colour characteristics or the overall colour perception of skin appearance. Existing studies have shown that, in terms of visual perception, the facial colour appearance could be

largely different from the colour appearance of other nonface objects. Considering the peculiarity of skin colour perception, how to accurately quantify and predict the perceived colour appearance of facial skin remains a challenge.

1.2 Aim and objectives

The aim of the present study is to understand the human colour perception of facial complexions, including colour preference and colour appearance. To achieve this aim, two objectives are set.

The first objective is to study the relationship between facial colour characteristics and preference judgements based on realistic skin models. The specific tasks include:

- To investigate the role of average skin colour in facial preference judgements (chapter 4).
- To investigate the role of various facial colour characteristics in preference judgements and identify their relative importance (chapter 5).
- To model the relationships between facial colour characteristics and preference judgements (chapter 6).
- To investigate the cultural difference between Caucasian and Chinese samples regarding facial colour preference (chapter 4, 5).

The second objective is to study the human perception of facial colour appearance. The specific tasks include:

- To quantify the overall facial colour appearance or the global colour impression of human faces (chapter 7).
- To understand the colour perception of facial whiteness, redness, and yellowness (chapter 8).
- To explore the perceptual difference between the colour appearance of the face stimuli and the nonface stimuli (chapter 8).

The approach taken in this study is psychophysics. Visual experiments have been conducted in preference evaluation and appearance assessments, all using images of real human faces.

1.3 Outline of the thesis

The flow chart below represents the research process of the present study. Nine chapters are included in this thesis. The overview of each chapter is also given below.
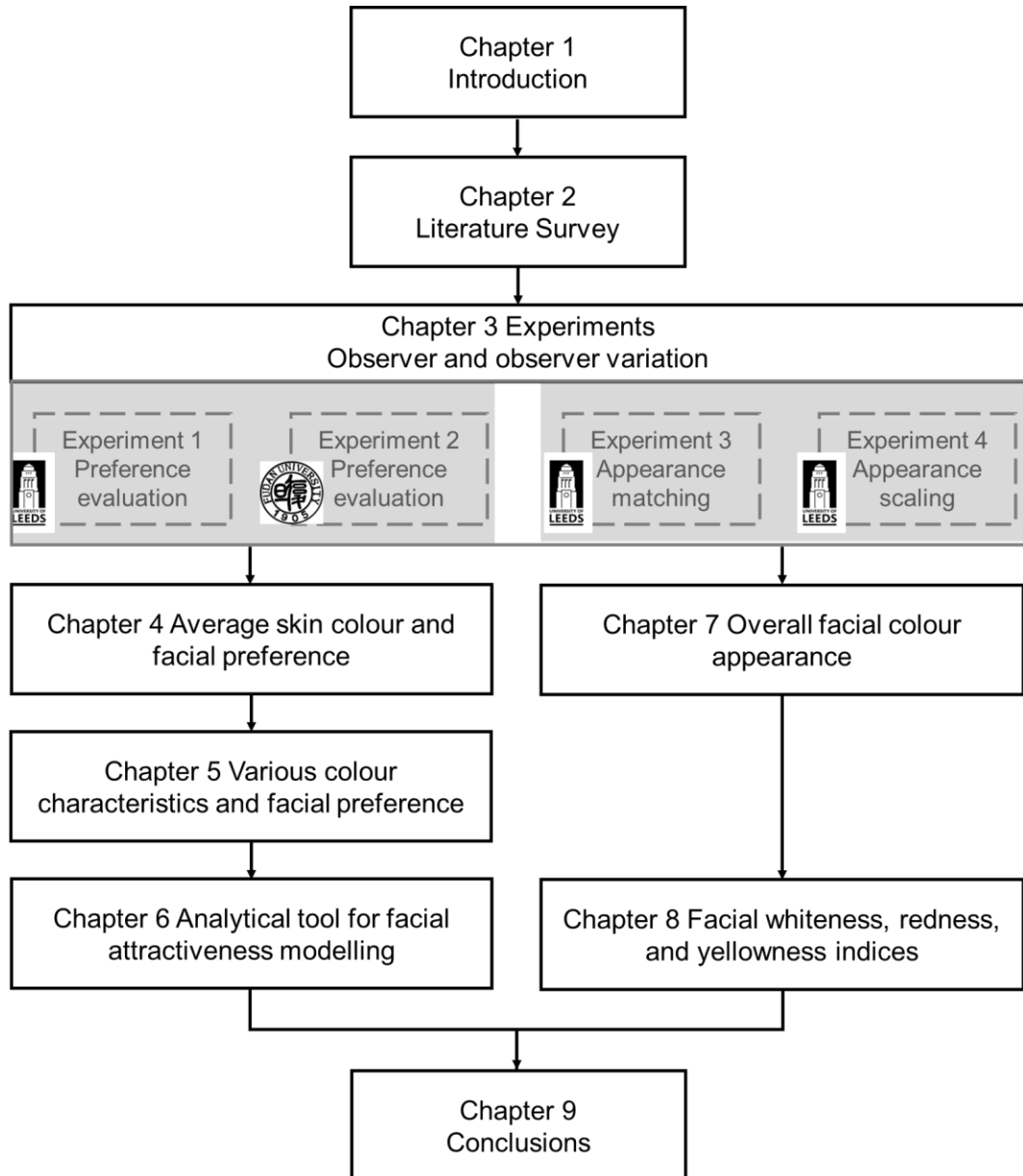
Figure 1.1 The structure of the thesis.

Chapter 1: Introduction

The current chapter introduces the general background, research aim and objectives of the present study, and the structure of the thesis.

Chapter 2: Literature survey

Chapter 2 gives a comprehensive literature review of the fundamentals of human colour perception, the CIE colorimetry system, research on facial

colour preference, research on facial colour appearance, display used to present colour stimuli, psychophysics, and data analysis techniques.

Chapter 3: Experiments

Chapter 3 focuses on the experimental preparations, including displays and images, and the details of the two groups of psychophysical experiments. The first group are two experiments related to facial colour preference, including Experiment 1, a cross-cultural experiment on preference evaluation conducted in Leeds, and Experiment 2, a partially repeating experiment of Experiment 1 conducted in Shanghai. The second group are two experiments related to facial colour appearance, including Experiment 3, a colour matching experiment on overall facial appearance, and Experiment 4, a colour scaling experiment on facial whiteness, redness, and yellowness. The information about observers and the observer variation is described in this chapter.

Chapter 4: Average skin colour and facial preference

Chapter 4 concerns the role of average facial skin colour in reference judgements. The average skin colour (L*, a*, b*) of eighty facial images was analysed and its relationship with three preference attributes, attractiveness, perceived healthiness, and perceived age was examined. The perceptual difference between the three facial attributes among Caucasian and Chinese observers was evaluated.

Chapter 5: Various colour characteristics and facial preference

Chapter 5 gives a comprehensive assessment of more facial colour characteristics, including the average facial colour, local skin colour, skin colour variation, and facial colour contrasts. Their role in facial preference judgements was tested and the relative importance was revealed. The cultural difference in the utilisation of various facial colour cues was revealed and discussed.

Chapter 6: Analytical tool for facial attractiveness modelling

Chapter 6 provides an analytical framework for facial attractiveness modelling from a large number of facial colour cues. The model performance was evaluated using both a training dataset and a novel testing dataset. Different regression techniques were compared based on their model performance and variable selection.

Chapter 7: Overall facial colour appearance

Chapter 7 is mainly concerned with the overall colour appearance of human faces. The colour perception of overall facial appearance was precisely examined and quantified. Factors that influence the perception of overall appearance were discussed based on several assumptions. The perceptual difference between Caucasian and Chinese observers was also discovered.

Chapter 8: Facial whiteness, redness, and yellowness indices

Chapter 8 concentrates on three colour appearance attributes of human faces, whiteness, redness, and yellowness. Their relationship with the CIELAB colorimetric values, L^*, a^*, and b^* were revealed and new indices were developed to accurately quantify and predict facial whiteness, redness, and yellowness. The perceptual difference in the three attributes between the face stimuli and the patch stimuli within the context of facial skin colour was also discussed.

Chapter 9: Conclusions

Chapter 9 summarises the main findings of the present study and discusses the directions for future work.

Chapter 2 Literature survey

2.1 Overview

This chapter provides the background information related to the present study. An outline of this chapter is given below.

Human colour perception (Section 2.2)

This section gives an overview of the fundamentals of human vision and colour perception, including the physiological aspects of the visual system, the basic perceptual attributes, and the colour appearance phenomena.

CIE colorimetry (Section 2.3)

CIE colorimetry serves as the foundation of colour specification and is an essential tool for colour quantification throughout this study. The CIE colorimetry system is introduced, and the uniform colour space, colour difference formulae, and colour appearance models are briefly reviewed.

Skin colour and facial preference judgements (Section 2.4)

This section reviews the research progress on facial colour preference. Using image-based methods, the effects of various colour cues, including the average skin colour and the other facial colour characteristics, on facial preference judgements are discussed. The limitation of the existing methods is considered.

Skin colour and facial appearance perception (Section 2.5)

After a review of facial colour preference, this section considers the quantification and the perception of the facial colour appearance. The methods of skin colour quantification are summarised. Studies on the perception of facial colour appearance and its perceptual difference from nonface objects are reviewed.

Display (Section 2.6)

This section introduces the characteristics and the colour characterisation process of the display, which is used as the medium for colour appearance assessment and preference evaluation in this study.

Psychophysics (Section 2.7)

Visual psychophysical experiments are the most important quantitative methods used in colour science research. This section introduces psychophysics and the psychophysical techniques used in this study.

Data analysis techniques (Section 2.8)

Finally, the statistical methods and modelling techniques used for data analysis in this study are introduced in this section.

2.2 Human colour perception

Colour is much more than a physical stimulus. It is detected by the human eye and interpreted in the brain. Human colour perception is rich and complex. As a result of the interaction between a light source, an object, and the eye and brain, or the vision system, colour perception also involves physiology, optics, neural processing, cognition, psychology, etc. (Berns, 2019). This section introduces the fundamentals of human vision and colour perception addressing some aspects of the human eye response and perceptual process.

Three books, *Measuring Colour* by Hunt and Pointer (Hunt and Pointer, 2011), *Foundation of Vision* by Wandell (Wandell, 1995), and *Colour Appearance Models* by Fairchild (Fairchild, 2013) are used as general references for this section.

2.2.1 Light and colour

Before understanding how the visual system processes the colour stimulus, it is necessary to characterise the nature of light which initiates the colour vision of human eyes. Light is the electromagnetic radiation that we can see, and it can be described by its wavelength in the unit of the nanometre (nm). As shown in Figure 2.1, the visible spectrum is limited between the wavelengths of 380 and 780 nm due to the sensitivity of human eyes. Shorter wavelengths have more energy than longer wavelengths. Note that the visible spectrum is continuous and there is no exact boundary between different colours (Hunt and Pointer, 2011). Figure 2.1 shows a rough correspondence between wavelengths and colours.

The Electromagnetic Spectrum
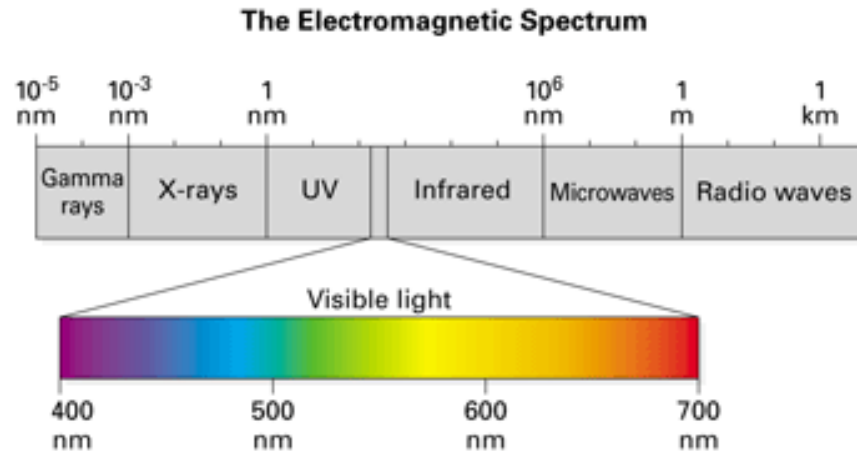
Figure 2.1 The electromagnetic spectrum and the visible light range (https://bodell.mtchs.org/OnlineBio/BIOCD/text/chapter8/concept8.2.html).

The early prismatic dispersion experiment conducted by Isaac Newton has shown that white light such as sunlight is polychromatic and can be separated into different colour components, whereas monochromatic light cannot be separated anymore (Newton and Hemming, 1704). In the real world, the majority of coloured stimuli are composed of many wavelengths (Berns, 2019). The changes in the wavelength components result in a change in the colour. The interaction between light and an object includes reflection, transmission, and absorption (Fairchild, 2013). Colour can be generated when the light is reflected or transmitted by a non-self-luminance object (e.g. an apple), or when the light is emitted by a self-luminance object (e.g. a display).

2.2.2 The visual system

The visual system includes the eyes, the connecting pathways through to the visual cortex and other parts of the brain, as shown in Figure 2.2. In this section, how the visual system responds to the physical stimulus, light, is introduced.
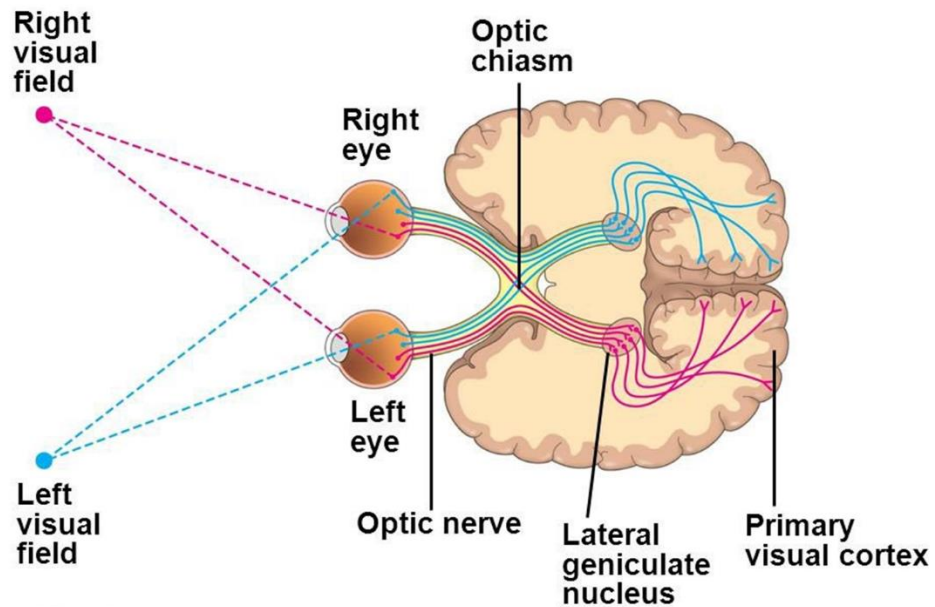
Figure 2.2 The human vision system (https://www.chegg.com/learn/biology/introduction-to-biology/neural-pathway-in-vision).

Figure 2.3 shows the schematically an eye with some key components. Light enters through the pupil. The cornea and lens act together as a compound lens to focus and project an inverted image onto the retina. The photoreceptors in the retina produce photochemical reactions and propagate nerve impulses through many layers of neurons to the brain and ultimately produce visual sensations.
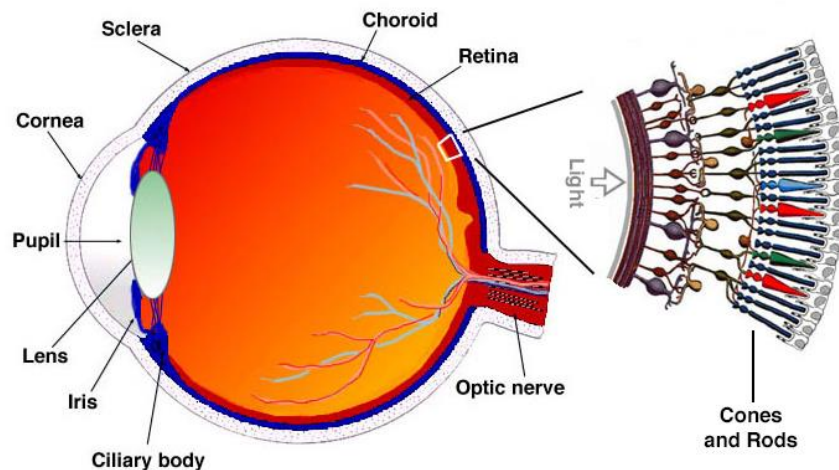
Figure 2.3 Cross-sectional diagram of the human eye and the retina (https://www.blueconemonochromacy.org/how-the-eye-functions).

There are two types of photoreceptors in the retina, rods, and cones. Rods gives monochromatic vision at low luminance levels (< 0.01 cd/m^2), which is

referred to as scotopic vision. Cones of three kinds give colour vision at normal luminance levels (> 10 cd/m^2), which is referred to as photopic vision. In the intermediate luminance levels, both rods and cones function and is referred to as mesopic vision. As a part of human vision, colour perception is mediated by the complex neural process starting with the stimulation of different photoreceptors in the retina.

Photoreceptors are not equally sensitive to light of all wavelengths. Human colour vision is served by three kinds of cones referred to as L, M, and S cones maximally responsive in long, middle, and short wavelengths, respectively. Figure 2.4 shows the spectral sensitivity curve of the L, M, and S cones, resulting in trichromatic colour vision.
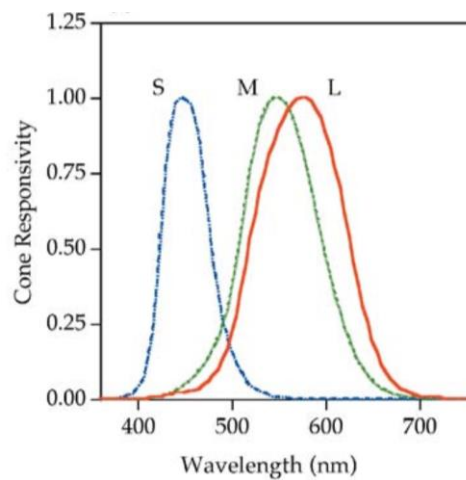
Figure 2.4 Spectral responsivities of the L, M, and S cones (Fairchild, 2013).

The image data obtained from the retina of the two eyes is first transduced into chemical and electrical signals in the photoreceptors, then processed through the retinal neurons network, the optic nerve formed by the ganglion cell axons, the lateral geniculate nucleus (LGN) in the thalamus, and to the visual cortex. About 30 visual areas have been in the cortex, labelled as V1, V2, V3, V4, etc. They are responsible for different aspects of the detection and interpretation of various visual information.

2.2.3 Mechanisms of colour vision

Historically, many scientists attempted to explain the mechanisms of colour vision. The two most convincing theories are the trichromatic theory and the opponent process theory. The trichromatic theory was proposed by Tomas Young in 1802 and extended by Hermann von Helmholtz in 1894. It assumes the retina's three types of cones are preferentially sensitive to the blue, green, and red colours respectively, and three images are formed by the three receptors and then transmitted to the brain (Young, 1845). The

opponent process theory was proposed by Ewald Hering. It suggests that the visual system interprets colour by opponent signals: red-green, yellow-blue and white-black (Hering, 1964). Both theories are empirically based and can explain various vision phenomena. The modern theory of colour vision tends to incorporate both theories into the stage theory which assumes a trichromatic response at the cone level (L, M and S cones) and an opponent colours response (luminance, red-green, and blue-yellow) in later stages. Figure 2.5 shows the colour vision model proposed by Vos and Walraven in 1971 (Vos and Walraven, 1971). The importance of the stage theory transforming from the trichromatic vision to the opponent processing is reflected in different colour appearance models (e.g. CIELAB).
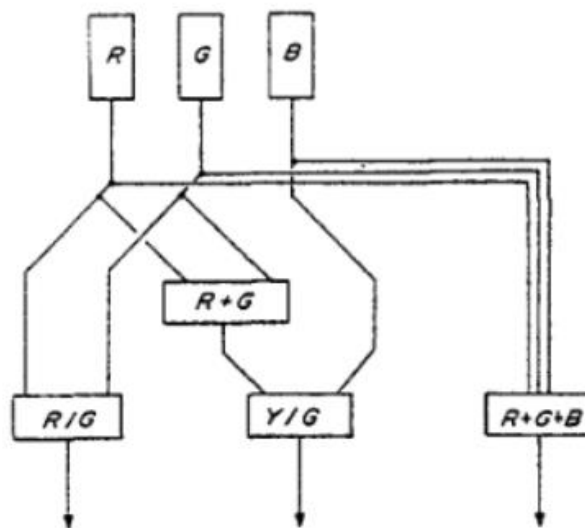
Figure 2.5 Colour vision model based on stage theory (Vos and Walraven, 1971).

Based on microspectroscopy and electrophysiology, the chromatic cone-opponency pathways in the LGN have been characterised and acknowledged. However, after the colour-opponent signals are sent to the visual cortex, the later cortical stages of visual processing are less well understood (Gegenfurtner, 2003). At the later stages, the encoding and processing of visual information become significantly more complex. Numerous ganglion cell responses are combined to produce various cortical responses and to further support visual capabilities such as colour perceptions, motion detections, etc (Fairchild, 2013). As the visual cortex and the brain involve, a higher level of colour perception and psychological cognition (e.g. preference, expectation) occur, which are difficult to explain from the perspective of physiology. Many important cognitive visual mechanisms could affect colour appearance, such as memory colour (the memory of the colour of familiar objects), colour constancy (the ability of

visual system to keep the colour of objects relatively stable despite the various illumination), but haven't been satisfactorily understood. For example, what is the role of the memory colour of familiar object in helping obtain the colour constancy (Granzier and Gegenfurtner, 2012). Based on the recent advances concentrating on the cortical processing of colour, the modern colour vision theory may see further progress in the future.

2.2.4 Perceptual attributes of colour

At the final stage of human colour perception, the visual stimuli of colour are not only considered as physical radiation. The appearance of a colour stimulus depends on the context in which it is seen. Colour appearance, representing the human perceptual attributes, is used to describe what colour stimuli look like under various viewing conditions at this stage. Several basic perceptual attributes which are used to specify the colour appearance are introduced in the section.

Three basic perceptual attributes of colour, brightness, hue and colourfulness are defined by Hunt and Pointer as below (Hunt and Pointer, 2011):

Brightness

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

Hue

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

Colourfulness

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

Three relative perceptual attributes of colours, lightness, and chroma are defined by Hunt and Pointer as below (Hunt and Pointer, 2011):

Lightness

The brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting.

Chroma

The colourfulness of an area judged in proportion to the brightness of a similarly illuminated area that appears to be white or highly transmitting.

Saturation

Colourfulness of an area judged in proportion to its brightness.

2.2.5 Colour appearance phenomena

Colour appearance phenomena, or substantive visual phenomena, describe the visual phenomena that affect how we perceive the appearance attributes of colour such as lightness, colourfulness, and hue (Elliot et al., 2015). In this section, details of a few colour appearance phenomena are given below.

Light and dark adaptation

Light adaptation describes the decrease in visual sensitivity when the level of illumination increases. An example is when we switch on the light in a dark room, our vision system is dazzled as it is overloaded due to the high sensitivity in the dark. After a short period of light adaptation, the sensitivity is decreased, and the vision goes back to normal. The dark adaptation refers to changes in the opposite direction and the process is slower than the light adaptation. An example is when we enter the dark movie theatre from the outside, it takes several minutes before we can see things clearly.

Chromatic adaptation

Chromatic adaptation refers to a far more important capability of the human visual system to adjust to the widely varying colour of illumination in order to approximately preserve the appearance of the object colour. It is the largely independent sensitivity regulation of the mechanisms of colour vision (Fairchild, 2013). Chromatic adaptation can be thought of as analogous to the feature of the automatic white balance of a camera. An example is that white paper appears to be white whenever viewed under daylight or candlelight.

Simultaneous Contrast

Simultaneous contrast causes colours to change in appearance when their background is changed (Albers, 2013). An example in Figure 2.6 shows that a black background causes the grey colour to appear lighter, whereas a white background causes the same grey colour to appear darker. The appearance change follows the opponent process theory, which implies that a darker background induces a lighter appearance; red induces green and green induces red; yellow induces blue and blue induces yellow; etc.
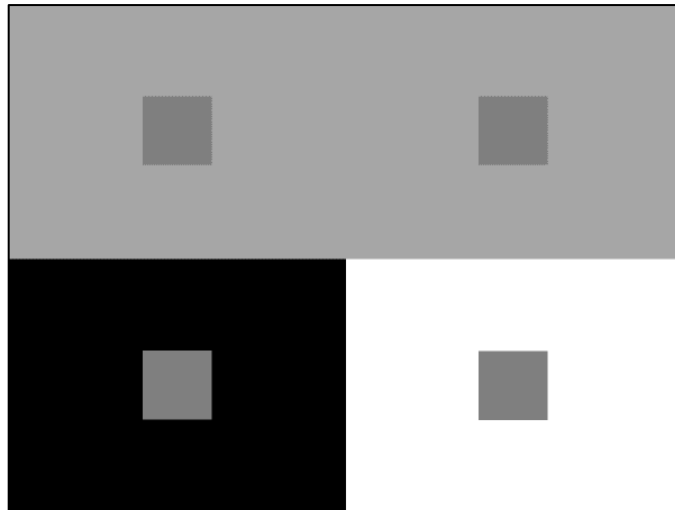
Figure 2.6 An example of Simultaneous Contrast, reproduced from (Fairchild, 2013).

Crispening

The increase in the perceived magnitude of colour differences when the background on which the stimuli are compared is similar to those of the stimuli (Craik, 1939). The effect can be illustrated in Figure 2.7, which shows the same two grey colours against a black, grey, and white background, respectively.
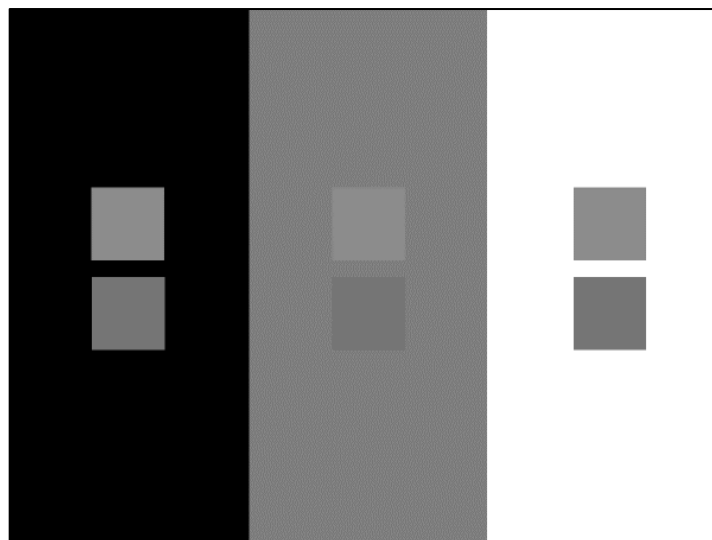
Figure 2.7 An example of Crispening, reproduced from (Fairchild, 2013).

Bezold-Brücke Hue Shift

Bezold-Brücke Hue Shift refers to a change in the hue of a colour produced by a change in luminance (within the range of photopic vision), while the chromaticity remains constant. Purdy's study has illustrated that wavelength change is needed to match the perceived hue of a monochromatic light at a

higher luminance level with another monochromatic stimulus at a lower luminance level (Purdy, 1931).

Abney Effect

Abney Effect refers to a perceived change in the hue of colour when the saturation is decreased, and the dominant wavelength and luminance remain constant (Burns et al., 1984). Thus, mixing a monochromatic light with white light does not preserve a constant hue.

Helmholtz-Kohlrausch Effect

Perceived brightness depends on both luminance and chromaticity. Helmholtz-Kohlrausch Effect shows that, at constant luminance, perceived brightness increases with increasing saturation, meanwhile the effect on brightness is also influenced by hue (Wyszecki, 1967). An example is shown in Figure 2.8. All the patches have the same luminance while some patches appear brighter than others.
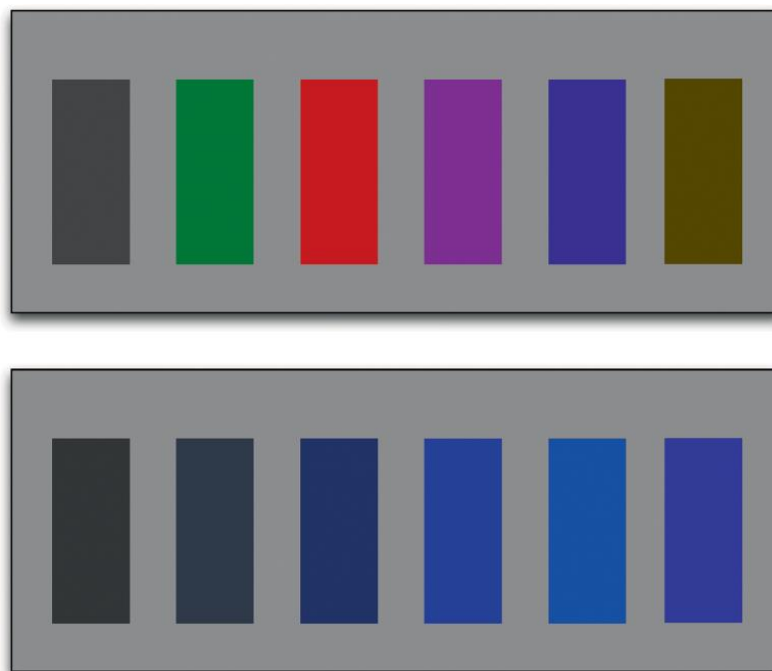
Figure 2.8 An example of the Helmholtz–Kohlrausch effect. All patches have the same relative luminance. Reproduced from (Elliot et al., 2015).

Hunt Effect and Stevens Effect

Hunt Effect refers to a progressive reduction in perceived colourfulness as the level of illumination falls (Hunt, 1952). Stevens Effect refers to a progressive reduction in brightness contrast as the level of illumination falls (Stevens and Stevens, 1963).

2.3   CIE colorimetry

Colorimetry is the science and technology used to physically quantify and describe human colour perception (Ohno, 2000). It serves as the foundation of colour specification and sets the stage for the development of colour appearance models (Fairchild, 2013). The Commission Internationale de l'Eclairage (International Committee on Illumination, CIE) is the primary organization responsible for the standardisation of colorimetry. The CIE colorimetry system numerically defined the three components required for a colour stimulus, the light source, the objects, and the human visual system, and quantified how they interact to produce colours. The system is used as an essential tool for the colour specification and further analysis throughout the current study.

In this section, *Colorimetry* by Ohta and Roberston (Ohta and Robertson, 2006), *Measuring Colour* by Hunt and Pointer (Hunt and Pointer, 2011), and *Colour Appearance Models* by Fairchild (Fairchild, 2013) are used as the general references.

2.3.1 Light source and CIE standard illuminants

Light sources such as candles, lamps, or sunlight, emit electromagnetic energy to initiate visual responses. They are characterised numerically by the spectral power distribution (SPD) curve, the distribution of energy at each wavelength across the visible spectrum (380 nm~780 nm). The SPD conventionally normalised to a value of 100 at the wavelength of 560nm is referred to as the relative spectral power distribution and is commonly used to describe a light source.

Another important characteristic of a light source is the correlated colour temperature (CCT), which is the temperature of the black body whose perceived colour most closely resembles that of the light source. For example, an incandescent lamp may have a CCT of 2800 K, a fluorescent tube of 5000 K, and an average daylight of 6500 K. Light sources perceive warmer (yellowish) with lower CCT and cooler (bluish) with higher CCT.

Light sources with different SPDs render the same object colour in different ways, which is inconvenient in quantitively expressing colours. Thus, the CIE established standard specifying the SPDs of three illuminants for use in colorimetry (ISO/CIE, 2022). They are CIE standard illuminant A, CIE standard illuminant D65, and CIE standard illuminant D50 (CIE illuminant D50 has recently been included as CIE standard illuminant, see ISO/CIE 11664-2:2022(E)). CIE standard illuminant A represents typical, domestic,

tungsten-filament lighting with a colour temperature of 2856 K. CIE standard illuminant D65 represents average daylight with a CCT of 6500 K. CIE standard illuminant D50 represents daylight with a correlated colour temperature of approximately 5000 K. Their relative spectral power distributions are plotted in Figure 2.9.
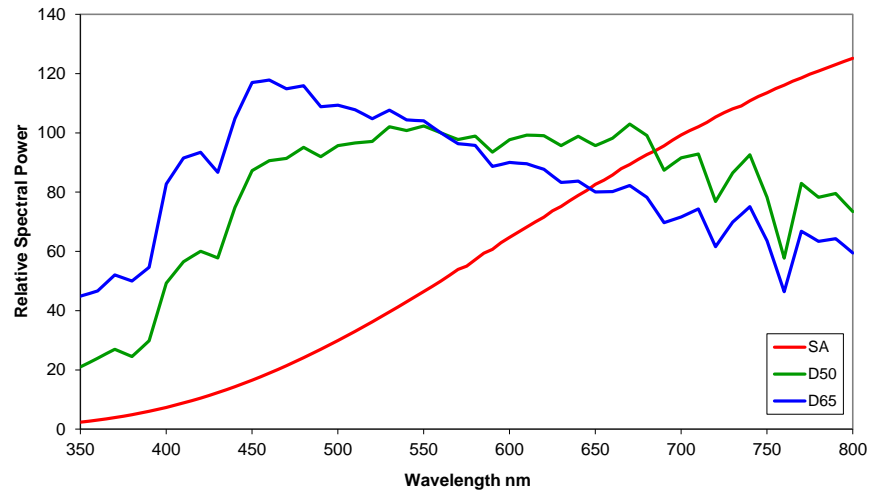
Figure 2.9 The relative spectral power distribution of CIE illuminants.

2.3.2 Colour-matching functions

The colour-matching functions are the numerical description of the chromatic response. Based on the trichromatic colour vision and Grassmann's laws of additive colour mixture, colour matching is processed by mixing the red, green, and blue light and controlling their amount to match a test light. Figure 2.10 shows a typical colour matching experiment. The three lights (typically red, green, and blue) required to match the test stimulus are referred to as primaries.
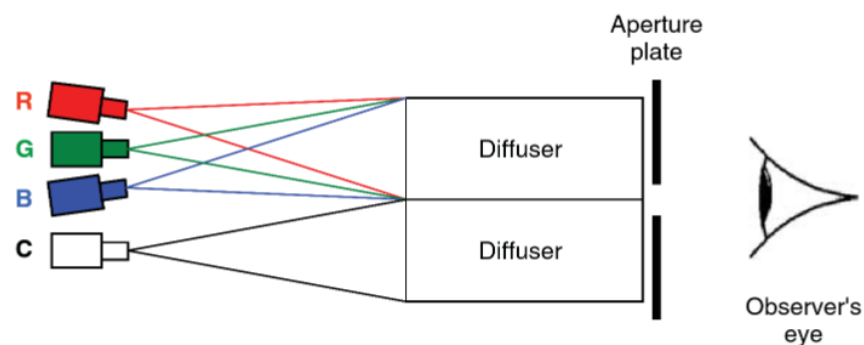
Figure 2.10 Principle of trichromatic colour matching by additive mixing of lights (Hunt and Pointer, 2011)

CIE defined colour-matching functions based on the two colour matching experiments conducted separately by Wright and Guild (Wright, 1929; Guild,

1931). Figure 2.11 (a) shows the colour-matching functions for the CIE 1931 Standard Colorimetric Observer, $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ represent the number of primaries (monochromatic stimuli of wavelengths 700 nm, 546.1 nm, and 435.8 nm) needed to match the stimulus at each wavelength. The curves were the average colour matching properties of 17 British observers, based on the experimental results of 10 observers obtained by Wright's investigation and 7 observers obtained by Guild's investigation.

A problem with the rgb colour-matching functions is that there are negative values of $\bar{r}(\lambda)$, which will add complexity to the calculation. To avoid the negative values, the primaries were then transformed from [R], [G], [B] to the unreal primaries [X], [Y], [Z] through a linear transformation. The transformation was carefully chosen so that X, Y, and Z would always be positive for all colours, and Y is proportional to L, the luminance of the colour specified. As the full-line curves in Figure 2.11 (b) show, this set of colour-matching functions is representative of the colour-matching properties of the CIE 1931 standard observer using approximately 2º field of view. In 1964, CIE has specified a supplementary set of colour-matching functions with a 10º field of view (the dash-line curves in Figure 2.11 b).
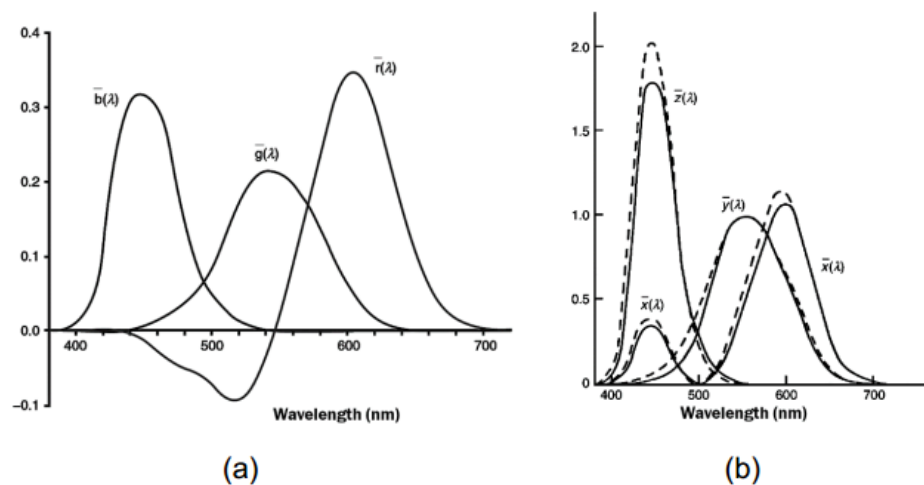
Figure 2.11 (a) The rgb colour-matching functions; (b) CIE 1931 xyz colour matching functions (full lines), and CIE 1964 xyz colour matching functions (dash lines) (Hunt and Pointer, 2011)

2.3.3 The tristimulus values XYZ

Considering the three components required for a colour stimulus mentioned above, the tristimulus values - CIE XYZ of the stimuli (non-self-luminous object) can be obtained:

$$X = k \sum_{\lambda} S(\lambda)R(\lambda)\bar{x}(\lambda)\Delta\lambda$$

$$Y = k \sum_{\lambda} S(\lambda)R(\lambda)\bar{y}(\lambda)\Delta\lambda \qquad \textbf{Equation 2.1}$$

$$Z = k \sum_{\lambda} S(\lambda)R(\lambda)\bar{z}(\lambda)\Delta\lambda$$

where $S(\lambda)$ is the spectral power distribution (SPD) of the light source; $R(\lambda)$ is the spectral reflectance of the object, also quantified as a function of wavelengths; $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are CIE colour-matching functions of the CIE 1931 standard observer as shown in Figure 2.11 (b); λ is the wavelength (in the unit of nm); k is a scaling constant to normalise the tristimulus values; k can be calculated by the following equation to make Y equal to 100 for the perfect diffuser:

$$k = 100 / \sum_{\lambda} S(\lambda)\bar{y}(\lambda)\Delta\lambda$$

Y then gives the luminance factor expressed as a percentage. The set of tristimulus values, XYZ, constitutes the units of CIE 1931 XYZ colour space and provides an objective description of colour sensations registered in the human eye.

To present the relative magnitudes of CIE XYZ tristimulus values, CIE chromaticity coordinates, x, y, z, are defined as the equations below:

$$x = \frac{X}{X+Y+Z}; \ y = \frac{Y}{X+Y+Z}; \ z = \frac{Z}{X+Y+Z} \qquad \textbf{Equation 2.2}$$

where x + y + z = 1. The two-dimensional chromaticity diagram (Figure 2.12) is widely used to illustrate colours using chromaticity coordinates, x, and y derived from the tristimulus values.
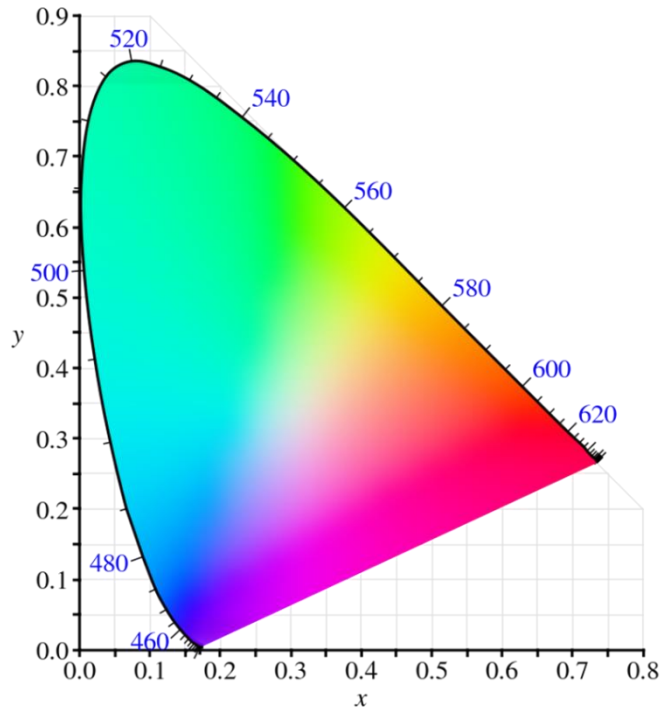
Figure 2.12 CIE 1931 x,y chromaticity diagram.

2.3.4 Uniform colour space

Though the XYZ colour space is very useful for colour quantification, an obvious limitation is that equal distance in the CIE xy chromaticity diagram doesn't have the perceptually equal colour difference. For example, the equal distance in the blue region has a larger colour difference than the green region perceptually. Thus, it is a non-uniform chromaticity diagram colour space. Besides, the chromaticity diagrams can only be used to compare colours with the same luminance. Considering both the chromaticity and luminance of colour, the two uniform colour spaces were developed, CIELAB and CIELUV, which provide uniform practices for the colour difference measurement. Both systems were recommended by the CIE in 1976.

The CIELAB uniform colour space (or CIE 1976 $L*a*b*$ uniform colour space) uses three orthogonal dimensions to describe the colour appearance of a stimulus, the vertical dimension $L*$ represents lightness, $a*$ represents the value along the red-green dimension, $b*$ represents the value along the yellow-blue dimension (Figure 2.13). Basically, in CIELAB uniform colour space, the same amount of numerical change of colour corresponds to the same amount of visually perceived colour change. The coordinates of CIELAB are transformed from the XYZ tristimulus values:

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

$$a^* = 500[f(X/Xn) - f(Y/Yn)]$$

$$b^* = 200[f(Y/Yn) - f(Z/Zn)]$$

Equation 2.3

$$f(w) = \begin{cases} (w)^{1/3} & for \ w > 0.008856 \\ 7.787(w) + 16/116 & for \ w \le 0.008856 \end{cases}$$

where X, Y, and Z and X_n, Y_n, and Z_n represent the tristimulus values of the object colour and reference white, respectively. The perceived attributes of chroma and hue can be predicted in CIELAB colour space. C_{ab}^* is obtained by the distance between the origin point and the colour point specified by the coordinates a^* and b^*:

$$C_{ab}^* = [(a^*)^2 + (b^*)^2]^{1/2}$$

Equation 2.4

Hue angle, h_{ab} is the angle used to specify hue expressed in positive degrees starting at the positive a* axis and progressing in a counter clockwise direction. h_{ab} is obtained by the following equation:

$$h_{ab} = tan^{-1}(b^*/a^*)$$ **Equation 2.5**
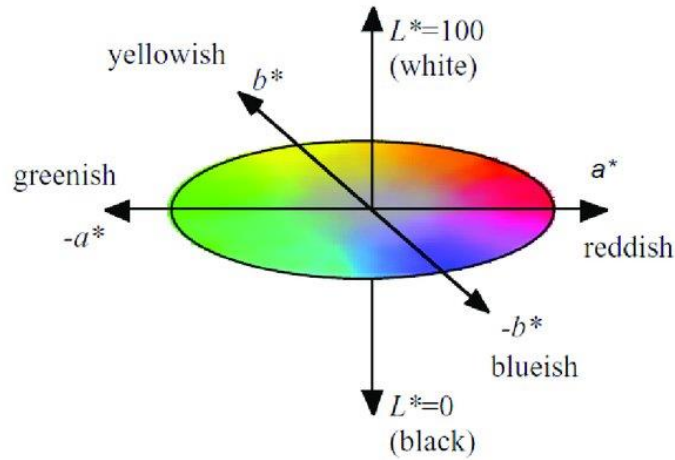
Figure 2.13 CIELAB uniform colour space.

The CIELUV uniform colour space (or CIE 1976 L*u*v* uniform colour space) also has three orthogonal dimensions, the vertical dimension L^* represents lightness, u^* represents the value along the red-green dimension, v^* represents the value along yellow-blue dimension:

$$L^* = 116(Y/Yn)^{1/3} - 16 \ \text{ for } \ Y/Yn > 0.008856$$

$$L^* = 903(Y/Yn) \ \text{ for } \ Y/Yn \le 0.008856$$

Equation 2.6

$$u^* = 13L^*(u'-u_n')$$

$$v^* = 13L^*(v'-v_n')$$

where u', v' and u'_n, v'_n are the chromaticity coordinates for the object colour and reference white, respectively, and are transformed from the XYZ tristimulus values:

$$u' = 4X/(X + 15Y + 3Z)$$

$$v' = 9Y/(X + 15Y + 3Z)$$

Equation 2.7

The CIELUV colour space correlates chroma, hue, and saturation in the following manner:

$$C^*_{uv} = [(u^*)^2 + (v^*)^2]^{1/2}$$

$$h_{uv} = tan^{-1}(v^*/u^*)$$

$$s_{uv} = 13[(u'\text{-}u'_n)^2 + (v'\text{-}v'_n)^2]^{1/2}$$

Equation 2.8

The CIELAB uniform colour space is widely used in the colorant and graphic arts industries, whereas the CIELUV uniform colour space is mainly used by the lighting, CRT and television industries. Moreover, the CIELUV uses substrative form of chromatic adaptation transform, which is extremely inaccurate in predicting corresponding colours data or colour difference (Fairchild, 2013). Thus, in the present study, the CIELAB system is used to describe the colour of non-self-luminous objects such as skin colour. Considering its uniformity and simplicity, it is used for colour specification, data analysis and model development.

Note that, the CIE colour spaces define colour independently of the device by which the colour is created or displayed, i.e. camera, display, printer, or scanner, thus is referred to as device-independent standard colour space. The device colour space, such as the sRGB space, is referred to as device-dependent colour space.

2.3.5 Colour difference formulae

Determining the difference between two colour stimuli is of great importance in colorimetry. Various colour difference formulae have been developed for quantifying such differences. The simplest forms of colour difference formulae, including ΔE^*_{ab} and ΔE^*_{uv}, are calculated by the Euclidean distance between the coordinates of two stimuli in the CIELAB and CIELUV colour spaces, respectively. Over the last thirty years, more complicated colour difference formulae have been proposed to more accurately predict the perceived colour difference, such as CMC (l:c), CIE94, CIEDE2000, etc. CMC (l:c) is a colour difference formula for small colour differences in the colorant industries (Clarke et al., 1984; McLaren, 1986). CIE94 and CIEDE2000 are elaborations of the CIELAB formula. As a replacement of

CIE94, CIEDE2000 is the current CIE recommendation for industrial colour difference prediction (CIE, 2018). It is applicable primarily for small colour differences between close patch samples (Sharma et al., 2005). It has limited performance for large colour difference prediction, and may not suit for assessing whole facial images. In this study, the CIELAB colour difference formula is used because it can accurately predict relative large colour differences and it is most used in previous studies to calculate skin colour difference (Stephen et al., 2011; Tan and Stephen, 2013; Melgosa et al., 2018; Amano et al., 2020).

The colour difference ΔE_{ab}^* is the Euclidean distance between two colours (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) in the CIELAB colour space, and it's obtained by:

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

Equation 2.9

Where

$$\Delta L^* = L_1^* - L_2^*$$
$$\Delta a^* = a_1^* - a_2^*$$
$$\Delta b^* = b_1^* - b_2^*$$

The colour difference ΔE_{ab}^* can also be determined by:

$$\Delta E_{ab}^* = [(\Delta L^*)^2 + (\Delta H_{ab}^*)^2 + (\Delta C_{ab}^*)^2]^{1/2}$$

Equation 2.10

Where

$$\Delta H_{ab}^* = 2(C_{ab,1}^* - C_{ab,2}^*)^{1/2} \sin(\Delta h_{ab}/2)$$

2.3.6 Colour appearance models

The CIE XYZ system specifies a colour stimulus in terms of the tristimulus values, but doesn't tell how it will appear. The appearance of colour stimuli is considered the final stage of human perception and depends on the viewing conditions, such as size, background, illumination, viewing geometry, etc. The same XYZ tristimulus values may lead to different visual perceptions when the colours are seen in different contexts. Thus, it is necessary to find a way of deriving the perceptual attributes of colours taking the influence of viewing conditions into consideration.

Colour appearance models are mathematical models extending tristimulus colourimetry toward the prediction of colour appearance under different viewing conditions. As defined by CIE Technical Committee 1-34 (Testing Colour Appearance Models), a colour appearance model is any model that includes predictors of at least the relative colour appearance attributes of

lightness, chroma, and hue (Fairchild, 1997). To produce predictors of these appearance attributes, a model must include a chromatic adaptation transform to predict corresponding colours across different illuminants (Fairchild, 2013).

Given the definition above, the CIE 1976 L*a*b* colour space (CIELAB) can be considered a colour appearance model. As the full CIELAB equations introduced in Section 2.3.4, the CIELAB colour space takes the XYZ tristimulus values and the reference white as input includes simple adaptation forms and produces predictors of lightness, chroma, and hue as output. Though CIELAB is a rudimentary colour appearance model, it is one of the most used models in various fields. Currently, in the area of skin research, the colour appearance is usually specified objectively in the CIELAB uniform colour space, thus it is adopted in the present study.

Considering the complexity of colour appearance phenomena as mentioned in Section 2.2.5, many attempts have been made to construct various more complicated colour appearance models for quantitively modelling human colour perceptions. Examples of colour appearance models proposed since CIE recommended CIELAB in 1976 include the Nayatani model (1990), the Hunt model (1991), the RLAB model (1996), the LLAB model (1996), the CIECAM97s model (1997), the CIECAM02 (2002), and the CIECAM16 (2016), etc. The CIECAM16 is the recent CIE recommendation for colour management systems (CIE, 2022).

Currently, all these existing colour appearance models were constructed using simple uniform colour patches. Based on the CIELAB colour space, the colour appearance of human complexions is studied in the current research and the perceptual difference between face and patch is considered.

2.4 Skin colour and facial preference judgements

Facial preference judgements have a profound impact on diverse important social outcomes, such as mate choices and social decision-making, thus it has been studied from various facial perspectives (Little et al., 2011; Rowland and Burriss, 2017). In particular, facial symmetry, averageness and sexual dimorphism have been widely studied over the years from an evolutionary or biological perspective (Thornhill and Gangestad, 1999; Rhodes, 2006). Compared to non-colour-related facial traits, the colour appearance of a human face has been relatively less investigated but has

gained increasing attention in the last ten to twenty years, which may suggest an important role for facial colour characteristics in any of the preference-related judgments including facial attractiveness, perceived healthiness, and perceived age. In this section, the studies on skin colour, including the average skin colour and the other facial colour characteristics, and its relationship with facial preference judgements are reviewed. The limitations of the widely used image manipulation method are summarised and the cultural difference is considered.

2.4.1 The average skin colour (L*, a*, and b*) and preference judgements

Facial preference judgements, including facial attractiveness, perceived healthiness, and visual age, have been studied with various facial colour characteristics being considered as individual crucial parameters. The average skin colour, including facial redness, yellowness, and lightness, was most widely studied and shown to play an important role in facial preference judgements. The majority of this research used L*, a*, and b* parameters in the CIELAB colour space to represent the facial lightness, redness, and yellowness, respectively. Table 2.1 summarises the literature studying the role of average skin colour in facial preference judgements.

Table 2.1 Studies on the average skin colour and facial preference judgements.

Authors	Ethnicity	IV	DV	Images	techniques
Stephen et al. 2009	Caucasian	L*, a*, b*	Health	manipulated faces	adjusting colour
Stephen et al. 2009	Caucasian	a*	Health	manipulated faces	adjusting colour
Re et al. 2011	Caucasian	a*	Attractiveness Health	manipulated faces	pair comparison
Stephen et al. 2012	Caucasian, African	L*, a*, b*	Attractiveness	manipulated faces	categorical judgement
Coetzee et al. 2014	Caucasian, African	L*, a*, b*	Attractiveness	manipulated faces	categorical judgement
Lefevre et al. 2015	Caucasian	L*, a*, b*	Attractiveness	manipulated faces	pair comparison
Pazda et al. 2016	Caucasian	a*	Attractiveness	manipulated faces	categorical judgement
Thorstenson et al. 2017	Caucasian	a*	Attractiveness Health	manipulated faces	categorical judgement
Foo et al. 2017	Caucasian	L*, a*, b*	Attractiveness Health	real faces	pair comparison

		L*, a*, b*, other biological indices			
Foo et al. 2017	Caucasian	L*, a*, b*, other biological indices	Attractiveness Health	real faces	categorical judgement
Appleton et al. 2018	Caucasian	b*	Health	real faces	categorical judgement
Tobitani et al. 2018	n.a. (skin patches)	melanin, haemoglobin	Attractiveness	manipulated patches	categorical judgement
Han et al. 2018	Caucasian, Chinese	L*, a*, b*	Attractiveness	manipulated faces	pair comparison
Jones et al. 2018	Caucasian	L*, a*, b*	Health	real faces	categorical judgement
Tan et al. 2019	Malaysian Chinese	L*, a*, b*	Attractiveness Health	manipulated faces	adjusting colour
Perrett et al. 2020	Caucasian	L*, a*, b*	Health	manipulated faces	pair comparison

IV = independent variable; DV = dependent variable.

Most studies used methods of image manipulation to conduct experiments. In these studies, observers were asked either to manipulate the facial colour to enhance their perceived preference (using the technique of adjusting colour) or to rate or make a forced choice between the colour-manipulated facial images in terms of their preference (using the techniques of categorical judgement or pair comparison). The images used in the experiments were normally computer-generated generic faces, which was usually morphed image averaged from several images of real faces. Figure 2.14 shows an example of manipulated facial images. As a result, increased facial skin lightness, redness and yellowness have been claimed to enhance the healthy appearance and facial attractiveness at a statistically significant level, mostly for Caucasian people.

Figure 2.14 Examples of face stimuli used in image manipulation studies (Thorstenson et al., 2017). Faces were manipulated on the CIELAB a* (redness) colour axis by -5 units (left) or +5 units (right).

For example, Stephen et al. investigated the role of overall skin colour in determining perceived facial healthiness by allowing participants to manipulate the skin portions of colour-calibrated Caucasian face photographs along CIELAB colour axes (Stephen, Coetzee, et al., 2009; Stephen, Law Smith, et al., 2009). In their results, to enhance healthy appearance, participants increased skin redness (a*) and skin yellowness (b*). Participants also increased skin lightness (L*), suggesting a role for low melanin coloration in the healthy appearance of faces. Lefevre and Perrett investigated the role of carotenoid colouration and melanin colouration in facial attractiveness judgements among Caucasian people (Lefevre and Perrett, 2015). They adjusted the level of these colourations by changing the colour of facial images along CIELAB colour axes (L*, a*, and b*) and had observers make attractiveness judgements on an internet-based pair-comparison test (thus the colour is not rigorously calibrated). They claimed that both increased carotenoid colouration and increased melanin colouration were found preferred compared to lower levels of these pigments and the carotenoid-linked health-signalling system was highly important in mate choices. Coetzee et al. studied African perceptions of female attractiveness and claimed that skin colour (lightness, yellowness and redness), skin homogeneity and facial adiposity significantly and independently predict attractiveness in female African faces (Coetzee et al., 2014). Re et al. conducted research to quantify the oxygenated blood colour (facial redness) change threshold required to affect the perception of attractiveness and health (Re et al., 2011). They found facial redness, to some extent reflecting the cardiovascular fitness of humans, had a perceptually equivalent influence on facial attractiveness and healthiness.

However, a few recent studies used non-manipulated real facial images for preference evaluation and found different results. For example, Foo et al. used images of real faces to study the predictors of facial attractiveness and health in Caucasian samples and revealed skin colour did not predict attractiveness in either sex and colour may play a limited role in determining attractiveness (Foo, Simmons, et al., 2017). Jones et al. investigated the influence of shape and colour cue classes on facial health perception and their results indicated that short-term health cues in the form of skin colouration showed no utilisation, with very weak correlations between perceived health and all three colour channels (L*, a*, and b*), all ps > 0.636 (Jones, 2018). Appleton et al. conducted a 4-week intervention in a randomized controlled trial where they documented a small but significant effect of fruit and vegetable intake on skin yellowness (about 1–2 b* units),

but no change in skin redness and no effect on perceived health (Appleton et al., 2018).

Considering the mixed results, it is not clear so far the role of the average skin colour in facial preference judgements, especially in real human faces. The two methodologies will be further summarised in Section 2.4.4.

2.4.2 Other facial colour characteristics and preference judgements

Apart from the average facial skin colour, other facial colour cues, including local skin colour, skin colour variation, and facial colour contrasts, have been found to influence facial preference evaluations. The related literature is summarised in Table 2.2.

The colour values at specific facial locations have been noticed by Jones et al. They defined three local facial areas and claimed cheek redness (a*) and periorbital luminance (L*) positively affected perceived health (Jones et al., 2016).

The appearance of the skin is noticeably uneven because of various structural details and colour variations, such as wrinkles, pores, spots, and freckles (Igarashi et al., 2007). The variation of skin colour, revealing information about skin texture, is also one of the facial colour characteristics that affect visual perceptions and preference judgements. Fink et al. conducted several studies to investigate the relationship between facial preferences and the homogeneity of skin colour/chromophore distribution. They used a set of shape-standardized stimulus faces with varying skin colour distributions and had observers rate these faces. They found that facial skin colour distribution significantly influences the perception and homogeneous skin colour distribution were perceived as younger and received significantly higher ratings for attractiveness and health than inhomogeneous skin colour distribution (Fink et al., 2006; Fink and Matts, 2008). Additionally, Fink et al. also used isolated cropped cheek patches instead of full facial images and demonstrated that the effect of skin colour variation on the perception of attractiveness, health, and age is independent of any facial shape or facial feature cues (Matts et al., 2007; Fink et al., 2011). Moreover, these perceptions of full-face images can be predicted by the isolated cropped skin images for all three attributes (Fink et al., 2012). Tan et al. also used cropped skin images to study perceived health in Chinese faces and similar results were found that homogenous skin texture linked positively apparent health of Chinese faces (Tan et al., 2018).

Facial colour contrasts also affect facial preference as cosmetics are widely used to alter facial contrast with the primary goal of making the face appears beautiful (Jones et al., 2015; Russell et al., 2019). Facial contrast, defined as 'the luminance and colour differences between the facial features and the skin surrounding those features' (Porcheron et al., 2013), is another cue for perceiving age (Porcheron et al., 2013; Porcheron et al., 2017), health ((Russell et al., 2016), and attractiveness (Russell, 2003). Most aspects of facial contrast were positively associated with attractiveness judgments, perceived health, and perceived age. In those studies, the adapted version of Michelson's contrast for three dimensions (L*, a*, b* coordinates in CIELAB colour space) between three facial features (eyes, eyebrows, and mouth) and their surrounding skin has been used to characterise facial contrasts (Michelson, 1995; Russell, 2009). Besides, Melgosa et al. studied the facial contrast differences between Caucasians and Orientals adopting the CIELAB colour differences ($\triangle$E) for facial contrast, and the CIELAB colour difference was found different between the two ethnic groups (Melgosa et al., 2018).

Table 2.2 Studies on other facial colour characteristics and facial preference judgements.

Authors	Ethnicity	Colour category	IV	DV	Images	techniques
Jones et al. 2016	Caucasian	local skin colour	forehead, periorbital, and cheeks L*, a*, b*	Health	manipulated faces	categorical judgement, pair comparison
Fink et al. 2006	Caucasian	skin colour variation	estimated age	Attractiveness Healthiness Age	manipulated faces	categorical judgement
Matts et al. 2007	Caucasian	skin colour variation	homogeneity algorithm	Attractiveness Healthiness Age	real patches	categorical judgement
Fink et al. 2008	Caucasian	skin colour variation	manipulated smoothness	Healthiness Age	manipulated faces	categorical judgement
Fink et al. 2011	Caucasian	skin colour variation	homogeneity algorithm	Attractiveness Healthiness Age	real patches	categorical judgement
Fink et al. 2012	Caucasian	skin colour variation	real age	Attractiveness Healthiness Age	real patches	categorical judgement
Stephen et al. 2010	Caucasian	facial colour contrast	Lip contrast - L*, a*, b*	Attractiveness	manipulated faces	adjusting colour
Russell et al. 2009	Caucasian	facial colour contrast	luminance contrast - L*	Attractiveness	manipulated faces	categorical judgement

Porcheron et al. 2013	Caucasian	facial colour contrast	colour contrast - L*, a*, b*	Age	manipulated faces	categorical judgement
Jones et al. 2015	Caucasian	facial colour contrast	colour contrast - L*, a*, b*	Attractiveness	manipulated faces	categorical judgement
Russell et al. 2016	Caucasian	facial colour contrast	colour contrast - L*, a*, b*	Health	manipulated faces	categorical judgement
Porcheron et al. 2017	Caucasian, Chinese, Latin American, South African	facial colour contrast	colour contrast - L*, a*, b*	Age	manipulated faces	pair comparison
Fink et al. 2001	Caucasian	average skin colour, skin colour variation	texture, colour	Attractiveness	manipulated faces	categorical judgement
Nkengne et al. 2008	Caucasian	average skin colour, skin colour variation	19 facial attributes (wrinkling, sagging, scaling colour, and texture)	Age	real faces	categorical judgement
Mayes et al. 2010	Chinese	average skin colour, skin colour variation	L*, a*, b*, other biological indices	Age	real faces	categorical judgement
Foo et al. 2017	Caucasian	average skin colour, skin colour variation	L*, a*, b*, other biological indices	Attractiveness Health	real faces	categorical judgement
Tan et al. 2018	Malaysian Chinese	average skin colour, skin colour variation	L*, a*, b*, Gabor factor A, B, C	Health (skin, face)	real faces and patches	categorical judgement
Russell et al. 2019	Caucasian	skin colour variation, facial colour contrast	with/without makeup	Age	manipulated faces	pair comparison

IV = independent variable; DV = dependent variable.

2.4.3 Facial colour characteristics were examined in isolation

The axiom in most previous research on facial preference described above is to change a single colour variable in a controlled experiment for preference evaluation. The single colour variable could be the average skin

colour, local skin colour, skin colour variation, or facial colour contrast. However, the impact of different colour cues taken together on facial preference judgment is not satisfactorily understood.

To test several cues together, the controlled experiment is not suitable anymore as it only consideres single variable, but the images of real faces or skin patches are needed for preference evaluation. A few notable exceptions are listed in the last six rows in Table 2.2. The exceptions include a study that compared average skin colour with structural facial features, which showed that skin colour did not predict facial attractiveness (Foo, Simmons, et al., 2017; Jones, 2018). Studies that investigated skin colour and various biophysical properties such as wrinkling and sagging on age perception, showed that skin colour had only a weak association with perceived age, while skin colour uniformity was the most important attribute (Nkengne et al., 2008; Mayes et al., 2010). Tan et al. used cropped cheek skin images to investigate the role of both skin colour and skin colour variation in health perception among Malaysian Chinese and claimed that homogenous skin texture and increased skin yellowness positively predicted the rated health (Tan et al., 2018).

Although these studies included more than one colour cue, the results are equivocal, and none considered all the different colour characteristics together (including the average facial skin colour, local skin colour, skin colour variation, and facial colour contrast). It is not known how these colour characteristics are taken together would affect facial preference, whether they have correlated themselves, and which characteristics are more important in terms of predicting facial preferences including attractiveness, healthiness, and visual age. Therefore, one aim of the present study is to investigate the effect of various colour characteristics on facial preference evaluation and compare their distributions in predicting facial preference.

2.4.4 The limitations of image manipulation

More importantly, the existing studies on the same colour predictors generated disputable results due to the different methodologies that were used. Concerning the widely used methods of image manipulation to provide the stimuli for the experiments, generally, much stronger associations between facial colour characteristics and preference have been revealed compared to recent studies using non-manipulated facial images. In studies that use image manipulation, statistically significant results have been found that increased facial skin lightness, redness and yellowness are linked to enhanced healthy appearance and facial attractiveness (Stephen, Coetzee,

et al., 2009; Stephen, Law Smith, et al., 2009; Stephen et al., 2012; Lefevre and Perrett, 2015; Pazda et al., 2016; Thorstenson et al., 2017).

To assess the role of various colour cues in facial preference judgment in real situations, the recently growing body of work has used realistic skin models in experiments for preference evaluation without any skin colour manipulation (Nkengne et al., 2008; Foo, Simmons, et al., 2017; Tan et al., 2018; Jones, 2018; Appleton et al., 2018). These studies, however, revealed very weak correlations between average skin colour and perceived healthiness ($p > 0.636$) (Jones, 2018), a limited role for colour in predicting attractiveness ($p > 0.05$) (Foo, Simmons, et al., 2017), and much weaker associations between skin colour and perceived age compared to skin colour uniformity or distribution (Nkengne et al., 2008; Mayes et al., 2010).

Although image manipulation could be an effective way to explore the effect of one single variable on preference evaluation while holding all other variables constant, it has several major limitations as summarised below:

- It may not be a reliable method to conduct comprehensive examinations of the various variables. With image manipulation, facial colour cues could only be studied individually, which may simplify the complex nature of facial preference judgement in real situations where various colour cues are considered together in a more holistical way.

- Using image manipulation, the role of the single colour characteristic that is being manipulated may be overestimated. Since observers can only manipulate a particular colour characteristic or choose manipulated facial images along fixed dimensions (e.g. CIELAB L*, a*, b*) for preference enhancement, they may only pay attention to that colour characteristic. Thus, the generated strong associations between the particular colour cue and the preference may not fit the real faces.

- Manipulated skin colour change could be impractical when uniform colour shifts are ideally applied to each pixel over the face. It is not necessarily consistent with naturally occurring coloration changes since the variation in the colour pattern depends on the distribution of blood vessels across the face which is not uniform.

- Manipulated skin colour changes are often restricted to a single dimension of CIELAB L*, a*, or b*. Real skin colour changes are not restricted to one of the CIELAB dimensions but are characterized by co-variations along all three dimensions, e.g. (Appleton et al., 2018).

- The manipulated skin colour could easily get out of the real skin colour gamut if it's not carefully processed. E.g. Figure 2.14 shows a magnitude of redness changes over 10 a* units and the colour shifts were applied uniformly across the face. Whereas in the real skin colour gamut, the range of redness values covered by different faces is about 6 a* units (see Figure 4.2).
- The computer-generated or morphed facial images may lose skin texture and appear to be unrealistic after image processing.

Considering the limitations above, the current study aimed to discuss facial colour preference within an evolutionary meaningful parameter space and to provide a useful and repeatable methodology for skin colour research based on a realistic skin model. High-resolution images of real human faces without changing the original colour were used, facial colour analysis was performed on each of the real facial images and a rigorous process of display colour characterization was performed to truly present the colour appearance of those facial images to observers in the preference evaluation experiments.

Using realistic skin models in experiments for preference evaluation could inevitably raise concerns that it is difficult for observers to ignore the role of certain facial features, for example, the eyes, the nose, the lips, and the mouth, and make judgements based only on skin colour. In the current study, this real variation is considered and is trying to be covered, at least partially, by using a large number of images of real faces. In fact, by using such a method, the results actually reflect the role of skin colour in facial preference judgments in a real situation where the preference could be more or less influenced by those structural facial features.

2.4.5 The cultural difference

Most research work that has investigated the impact of facial colour appearance on these perceived attributes was conducted using Caucasian samples, both as participants and to provide stimulus material (see the column 'Ethnicity' in Table 2.1, Table 2.2). Facial colour perception, however, may vary between different ethnic groups. In colour imaging, people typically rely on their perceptions and preferences to judge the quality of the colour reproduction of faces. It is important to know the perceptual differences for preferred colour reproduction when diverse populations of both observers and the observed are involved. Cultural differences in the perception of facial colour appearance are also worthy of consideration because of the increasing number of applications that need to define the preferred colour reproduction based on the needs of different people,

including photography and graphic arts, dermatological diagnosis and surgery in medical applications, mannequin display in retail and e-commerce, the product development of cosmetics, and colour rendition under various types of light source (Zeng and Luo, 2010; Xiao et al., 2014; Okuda and Okajima, 2017; Gao et al., 2018).

So far, four studies have focused on the perceptual cultural differences in facial colour preference. Two studies by Stephen et al. were conducted amongst Caucasian and African populations(Stephen et al., 2012; Coetzee et al., 2014). In the first, both Caucasian and African observers viewed facial images of their own ethnicity and found similar perceptual preferences for increased skin lightness (L*) and yellowness (b*) in both Caucasian and South African populations. The second study also found that an association between skin colour and male facial attractiveness only existed when viewing own-ethnicity faces, both for African and Caucasian observers. A study conducted by Han et al., however, did not find a cross-cultural similarity in facial colour preference but found significantly different preferences for facial colour between Chinese and Caucasian participants such that Chinese observers prefer lighter skin and decreased yellowness compared to Caucasian participants (Han et al., 2018). Malaysian Chinese, by contrast, linked increased yellowness and redness but decreased lightness with enhanced perceived healthiness (Tan and Stephen, 2019). Porcheron et al. studied the influence of facial colour contrast on age perception among Caucasian, Chinese, Latin American, and South African (Porcheron et al., 2017). They revealed that facial colour contrast was a cross-culturally valid cue for perceiving age and that increasing the facial colour contrasts made the faces look younger, which was independent of the ethnic origin of both faces and observers.

The limited cross-cultural studies inclusively used methods of image manipulation and showed controversial results. It remains unclear whether the perception of facial attractiveness, healthiness, and age, follows a similar pattern for different ethnic groups, and whether cultural differences exist in the preference judgements of real human faces. In the current study, the cultural difference in facial preference judgement was investigated between Caucasian and Chinese populations.

2.5 Skin colour and facial appearance perception

The great research interest in skin colour and facial preference judgements has shown the importance of facial colour appearance. The facial colour

appearance describes what the colour stimuli of facial skin look like in human colour vision. Although CIE colorimetry provides the objective tool for colour measurement and quantification, colour perception is subjective in nature, and the perceptual aspects are important in any evaluation of human faces. In this section, the methods of skin colour quantification are summarised. More importantly, studies on the perception of facial colour appearance and its perceptual difference from nonface objects are reviewed.

2.5.1 Skin colour quantification

It is critical to describe skin colour characteristics accurately through effective measurements in skin colour research. Human skin is a complex multi-layered surface with colour unevenly distributed. These properties add to the uncertainties of accurate colour measurements. Several types of measuring methods have been used for quantifying skin colour. The three most commonly used methods are described in this section, and they are colour charts, instrumental measurements, and image-based measurements.

Colour charts

Colour charts are used to classify human skin colour through colour matching with the target skin. The method has been adopted a long time ago before measuring instruments were introduced and is still used today. Table 2.3 summarises five colour charts used for skin classification with various needs (Fitzpatrick, 1988; Taylor et al., 2005; De Rigal et al., 2007; Swiatoniowski et al., 2013).

Table 2.3 Colour charts for skin classification.

Colour charts	Authors	Descriptions	Purpose	Samples
Von Luschan's chromatic scale	Felix von Luschan late 1800s	36 coloured opaque glass tiles	Measure skin colour, evaluate skin treatments	
Fitzpatrick skin prototype scale	Fitzpatrick 1975	6 skin colour types	Classify skin colour for UV protection	

Taylor hyperpigmentation scale	Taylor, S. et al. 2006	15 coloured plastic cards	Evaluate hyperpigmentation treatment	
Skin Colour Chart®	L'Oréal® 2007	52 fan charts	Assist clinicians in evaluating the skin care products	
Pantone SkinTone™ Guide	Sephora and Pantone	110 fan charts	Personalised cosmetic shopping	

Colour charts serve as an inexpensive and convenient tool to assess skin tone and thus have been successfully used in cosmetic companies. However, the accuracy or consistency of colour measurements largely depends on visual assessments of individuals and viewing conditions. Besides, colour charts can only have a limited number of colours and are not enough to determine human skin colour with a wider range and more variation. These limitations could be overcome by instrumental measurements and image-based measurements.

Instrumental measurements

Point measuring methods by instruments have also been most widely adopted for skin colour measurement. Tele-spectroradiometers (TSR) and spectrophotometers (SP) are two types of instruments that are widely used to acquire skin colour and spectral information. Spectroradiometers is a non-contact-type instrument and require an external light source to measure the spectral power. Spectrophotometers is a contact-type instruments with a built-in light source and can measure the spectral reflectance of the skin surface. Based on CIE colourimetry, the skin colour of the target point area can be obtained in terms of the CIE tristimulus values and CIELAB coordinates, L*, a*, and b* values. Figure 2.15 shows an example of using a spectrophotometer to measure the skin colour of a subject's forehead.
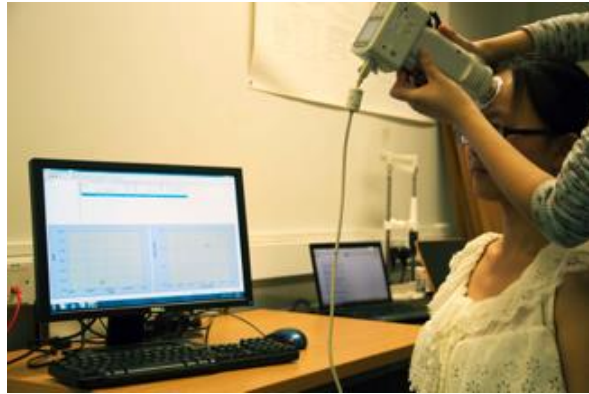
Figure 2.15 Example of using a spectrophotometer to obtain measurements on the subject's forehead (Wang et al., 2018).

Both instruments are more sensitive and reliable than visual assessments, thus can give more accurate point measurements of skin colour. Actually, the Skin Colour Chart® developed by L'Oréal® was based on the skin data measured by spectroradiometer instruments and the Pantone SkinTone™ Guide was established based on the spectrophotometer measurements. Although instruments can give relatively precise colour measurements, some uncertainties should be taken into consideration while using them. For example, the repeatability of the non-contact spectroradiometers is affected by measurement distances, and the measuring results of the contact spectrophotometers are affected by both measurement field sizes and different pressure applied during measurements (Wang et al., 2018). Furthermore, the instruments can only be used to obtain the colour information of certain points, but cannot include the overall skin colour information and all the colour characteristics, e.g. the average facial skin colour, skin colour variations, facial colour contrasts, etc.

Image-based measurements

Obviously, the point measurements are not enough to acquire the various facial colour characteristics mentioned in Section 2.4. In this regard, image analysis using digital cameras provides another method to record and assess skin colours. The image-based methods for colour measurements have been a research interest because they can be used to acquire a huge amount of information about the whole image and meanwhile achieve accurate colour measurements (Miyamoto et al., 2002; Xiao et al., 2016; Kikuchi et al., 2020). For the camera imaging systems, a stable environment with defined lighting conditions is required and mathematical algorithms are needed to transform the digital image data to CIE XYZ tristimulus values or spectral reflectance values. Thus, the colour specifications of each pixel in the facial image can be obtained. Moreover, the average facial skin colour

and other facial colour characteristics, such as skin colour variation and facial colour contrasts can be calculated based on the image data. Facial images are quite essential for skin colour presentation and facial attractiveness evaluation, so they have been used frequently in relevant studies (as reviewed in Section 2.4).

Based on instrumental measurements and image-based measurements, several international skin colour databases have been established by different researchers to meet multi-disciplinary applications. Skin colour database collects skin colour information, investigates the variation in skin colour between different ethnic groups, gender and body locations, and provides a vital reference for skin colour research. In 2003, the International Organisation for Standardization published an international standard: Graphic technology - Standard object colour spectra database for colour reproduction evaluation (SOCS) (Tajima et al., 2002). SOCS recommended six sets of skin colour, SHISEIDO, KAO, OOKA, KAWASAKI, OULU, SUN (Tajima et al., 1998; Marszalec et al., 2000; Sun and Fairchild, 2002). The first three databases have not published their measurement setups. The latter three databases captured facial images during data collection, but all the facial images were not included in the published database. Besides, a Chinese skin colour database was built in 2012 with 202 Chinese subjects (Xiao et al., 2012).

A more comprehensive skin colour database, the Liverpool-Leeds Skin-colour Database (LLSD), with unified measurement methods and covering different ethnic groups, genders, ages and body locations was then established by the Universities of Liverpool and the Universities of Leeds (Xiao et al., 2017; Wang et al., 2018). Figure 2.16 shows how the facial image was captured in a diffused lighting condition, and the digital SLR camera was positioned in front of the subject's face. This database included data for 188 subjects from four ethnic groups (Caucasian, Oriental, South Asian and African, including both genders). The facial images captured by a digital camera under CIE illuminant D65 and the reflectance data of each pixel in the images were also included in the database. The variations in ethnic skin colours were investigated based on the database (Xiao et al., 2017).

In the current study, the facial images selected from LLSD are used to assess the skin colour appearance and facial attractiveness in all the experiments. The process of image selection and processing will be described in Section 3.2.2. Based on these facial images of real faces,

accurate colour information is obtained to support the following data analysis and modelling work.
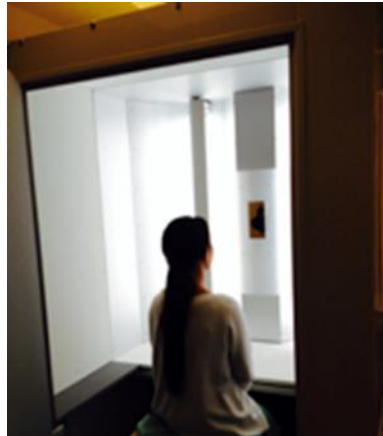
Figure 2.16 Example of facial image capturing using the using VeriVide DigiEye® light booth (Wang et al., 2018).

2.5.2 The overall facial colour appearance

As reviewed in Section 2.4, the impact of facial colour appearance on preference judgements has been widely studied using the image-based method. Those studies have commonly used the mean pixel colour of facial skin area specified in CIELAB colour space (average L*, a*, and b* values) to represent the overall facial colour appearance. Although the facial image contains colour information of the entire face and image-based measurement can achieve high accuracy, the human colour perception could be different from the colour measurement. The overall colour appearance of human faces has not been studied so far and whether the overall facial colour appearance is perceived as the same as the mean pixel colour remains unknown.

In real life and many applications, the colour appearance of non-uniform surfaces, such as teeth, food, and textiles, is evaluated. The human visual system seems to be able to extract and summarise the representative colour description, e.g. tree foliage is green and banana is yellow (Kuriki, 2004). Yet little is known about the formation of the summary representations of multi-coloured stimuli. The conventional hypothesis is the colorimetric average hypothesis that the colour appearance, or say a single colour impression, is determined by the colorimetric average of the elemental colours that constitute the textured pattern. However, as the colour appearance of the multi-coloured patch and several objects have been studied in a small number of literature, the human colour perception of multi-

coloured stimuli seems cannot be simply explained by the colorimetric averages.

Sunaga and Yamashita studied the global colour impressions of the non-uniform textured patch consisting of two colours with the same hue and brightness but different saturation (Sunaga and Yamashita, 2007). Using an asymmetrical colour matching experiment, they found the matched colour shifted toward a higher saturation than the colorimetric averages, which supported the colour appearance hypothesis that the single colour impression is determined by an integration of the appearance of the elemental colours. Giesel and Gegenfurtner studied the colour appearance of real objects varying in material, hue, and shape (Giesel and Gegenfurtner, 2010). In their experiments, observers adjusted the colour of the uniform patch to match the colour appearance of the object and results also showed observers did not simply take the average colour across objects. Rather, the geometry of the objects was taken into account and the variations in reflected light were neglected. Kimura investigated how colour information was summarized in multicolour mosaics by matching a uniform colour patch to the multicolour mosaics (Kimura, 2018). It was found when the colour variation was large, the matched colour deviated from the colorimetric mean toward the most-saturated colour. Virtanen et al. conducted experiments to characterize different aspects of the spatial integration of hue and claimed humans are efficient at integrating hue information over space (Virtanen et al., 2020).
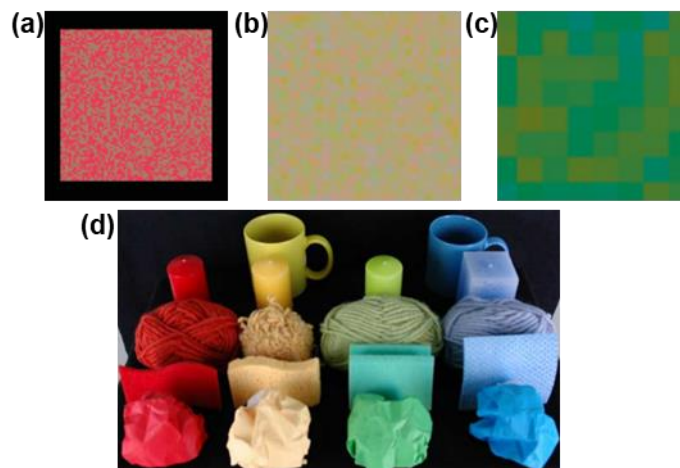
Figure 2.17 The multi-coloured stimuli used in previous studies on overall colour appearance (a) (Sunaga and Yamashita, 2007), (b) (Kimura, 2018), (c) (Virtanen et al., 2020), (d) (Giesel and Gegenfurtner, 2010).

So far, the studies on overall colour appearance or global colour representations of multi-coloured stimuli are restricted to relatively simple

objects, including textured patches, mosaics, monochromatic objects, etc. (as shown in Figure 2.17). Whereas human faces are more complex visual stimuli with larger variations and more unique patterns. How the visual system discriminates, identifies, integrates the colour information, and forms ensemble perceptions across the whole face could be quite different and complicated. In the current study, how the colour appearance of a human face is summarised as a representative colour is studied. The relationship between the overall colour perception and colorimetric average values of human faces is investigated.

2.5.3 Facial whiteness, redness, and yellowness

Facial skin whiteness, redness and yellowness are three attributes most directly describe people's perceptions of facial colour appearance and thus receive most concerns in application fields such as cosmetics and dermatology and also in our daily life. Studies on facial colour and preference judgements have commonly used the CIELAB colour space to describe skin colour appearance, and the overall facial lightness, redness and yellowness have been simply represented by the mean L*, a* and b* values of the facial skin area respectively (Section 2.4.1). So far, not only the overall facial colour appearance is not clear, but the perception of facial redness and yellowness has also not been precisely examined. It is unknown whether these colorimetric values in CIELAB colour space are equivalent to the subjective colour perceptions of facial skin.

Whiteness is an important colour attribute in many industries and thus indices have been developed for evaluating the whiteness in different applications, most notably for paper and textiles. For example, the CIE whiteness index (WIC) of neutral hue was a widely used index recommended by CIE (ASTM, 1993). Besides, WIO and WI_D were specifically used to assess the tooth whiteness perception in dentistry (Luo et al., 2007; Luo et al., 2009; Pérez et al., 2016). Those indices are limited to a specific range of chroma and tint (nearly white). To describe general colours, the depth scale D^*_{ab} was developed by Berns to describe the difference in colour from the neutral colours (Berns, 2014). It is calculated by the distance between the whitest colour (L*=100, C^*_{ab}=0) and the sample colour in the L* C^*_{ab} plane in the CIELAB colour space. In the area of skin colour, whiteness or having white skin is considered an important element in female beauty in Asian cultures (Xie and Zhang, 2013; Gao et al., 2018). The Individual Typology Angle (ITA°) was proposed to classify skin colour with 'lightness' and show the degree of constitutive pigmentation of the skin

(Chardon et al., 1991; Del Bino and Bernerd, 2013). The term 'lightness' used here is not the L* in the CIELAB system, but similar to the meaning of skin whiteness. As Figure 2.18 shows, the ITA° index is determined in the L*b* plane based on the CIELAB system and allows skin colour types to be classified into six groups. For the skin colour with lightness higher than 50, the skin colour with higher L* and smaller b* values appears lighter. For the skin colour with lightness lower than 50, the skin colour with higher L* and bigger b* appears lighter. This whiteness scale indicates the difference between perceived skin whiteness and colorimetric values of lightness (L*). On the other hand, studies on brightness perception showed though perceived facial whiteness was found highly associated with lightness (L*) as well as brightness (Shimakura and Sakata, 2019; He et al., 2021), whiteness is perceptually independent from brightness in face perception (Shimakura and Sakata, 2019). Another study on facial skin whiteness conducted among Japanese women has found that whiteness is not only linked with L* and b* but also related to a* (Yoshikawa et al., 2012). They claimed that reddish facial skin appeared whiter than yellowish one in high-lightness regions and low-chroma facial skin appeared whiter than high-chroma one.
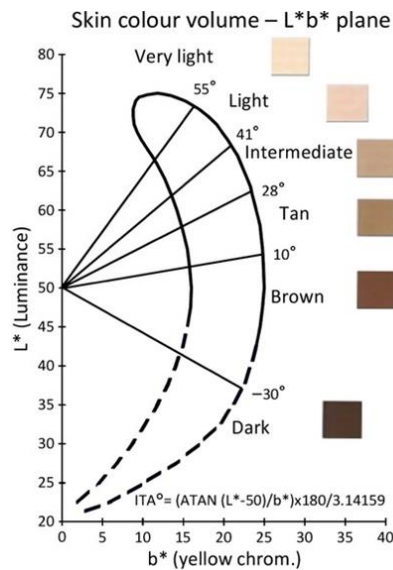
Figure 2.18 The Individual Typology Angle (ITA°) (Del Bino and Bernerd, 2013).

The Yellowness index has also been developed in dentistry (Sullivan et al., 2019), whereas no index of redness or yellowness could be used to quantify the colour perception of human complexion. Facial skin redness is determined by perfusion with oxygenated blood and conveys information about cardiovascular fitness and emotional state (Re et al., 2011). Facial redness is also considered as the most common and recognisable clinical sign of many facial dermatoses (Dessinioti and Antoniou, 2017). Skin

yellowness is influenced both by dietary carotenoids and melanin (Lefevre and Perrett, 2015). Interestingly, facial redness and yellowness perceptions have shown special salience in colour change discrmination compared to facial luminance (Tan and Stephen, 2013). Both facial redness and yellowness were found closely related to facial attractiveness, healthiness, and visual age (Section 2.4.1). However, those studies simply used a* and b* parameters of the CIELAB colour space to roughly characterise facial skin redness and yellowness, respectively. Though in the CIELAB system, a* represents the value along the red-green dimension, and b* represents the value along the yellow-blue dimension, we haven't know whether a* and b* are accurate measures of the subjective colour perceptions of 'skin redness' or 'skin yellowness'.

The literature above suggested that facial skin whiteness might not be influenced only by L*; though skin redness and yellowness were expected to be connected with a* and b*, respectively, it was far from clear what the actual relationship is within the constraints of skin colour space. Considering the peculiarity of human skin colour, skin whiteness, redness, and yellowness might be good scales to quantify facial colour appearance from the perspective of visual perception. In the current study, psychophysical data is collected to quantitatively define these perceptual attributes of skin colour appearance and to precisely examine their relationship with the colorimetric values.

2.5.4 Perception of facial colour appearance

Perception of facial colour appearance is unique. The existing studies have shown that, in terms of colour discrimination and perception, the facial colour appearance is largely different from the colour appearance of nonface objects.

Concerning facial colour discrimination, Tan and Stephen examined the detection threshold for skin colour changes in Malaysian Chinese observers (Tan and Stephen, 2013). They revealed visual sensitivity was greater at discriminating changes in facial redness and yellowness than changes in luminance. However, the sensitivity was not greater for nonface stimuli (colour patches). Meanwhile, they also found the cross-race effect in discriminating facial colours that participants were more sensitive in recognising colour changes of their own faces rather than faces of other ethnicity.

Regarding colour appearance perception, a few studies were conducted comparing the face stimuli and the uniform patch stimuli. Yoshikawa et al. studied the influence of chromatic components on whiteness perception and found that perceived facial whiteness is largely influenced by chroma and hue that reddish skin and low-chroma skin appeared brighter or whiter (Yoshikawa et al., 2012). Yet these results were not found in the perception of the uniform colour patch, which suggested a higher-level process of face recognition existed and affected the perception of skin whiteness. Shimakura and Sakata investigated the effect of saturation on perceived brightness and whiteness in both face images and uniform colour patches (Shimakura and Sakata, 2019). They found an inverse effect on the perception of the colour appearance of the face and uniform colour patch. A uniform colour patch appeared brighter with increased saturation (the Helmholtz–Kohlrausch effect) while a facial image appeared less bright with increased saturation while (contrary to the Helmholtz–Kohlrausch effect). They also suggested the contribution of higher-order recognition mechanisms. Hasantash et al. tested how memory modulated the colour appearance perception of objects and faces under different light conditions and found, in some lighting conditions, the impact of memory on colour perception is greatest for face colour compared to other colours (Hasantash et al., 2019). More recently, He et al. investigated the effect of colour on brightness perception among different Asian countries and suggested the perceptual difference existed among different countries (He et al., 2021). Normally such cultural dependency would not be expected in the perception of uniform colour samples.

All the literature above has noted the distinctiveness of facial colour perception from nonface objects. In the current study, both the overall facial appearance and the perception of skin whiteness, redness, and yellowness are assessed. Meanwhile, the perceptual difference between the face stimuli and the uniform patch stimuli with the same colour appearance is explored with the aim of better understanding human perception of skin colour.

2.6 Display

Two displays, including an LCD and a LED display, were used in this study as a medium for colour appearance assessment and preference evaluation. It is of key importance to reproduce the colour of the visual stimuli (images of real human faces and uniform colour patches) accurately and consistently on display during the psychophysical experiments.

2.6.1 Display calibration and characterisation

The display colour is determined by the RGB digital signals. As they are device-dependent, different displays are likely to have different RGB primaries and generate different colour appearances of the same image. The same display is also likely to generate different colour appearances depending on the settings (e.g. brightness, contrast, sharpness) and status of the device (e.g. warm-up time) (Westland et al., 2012). To solve this problem, display calibration and characterisation were conducted. As defined by Johnson (Johnson, 1996), calibration is the process of setting up the device to ensure it gives consistent responses every time; characterisation is the process of defining the relationship between the device-dependent colour space such as the adobe RGB space and the device-independent colour space such as the CIE XYZ. In the following section, the display characterisation models will be introduced. The settings under which display characterisation is conducted should be kept the same all the time during the psychophysical experiments to ensure the displays give a consistent performance.

2.6.2 Display characterisation models

Normally, display characterisation models can be simplified into a two-stage model. The first stage is a non-linear transformation between the RGB signals and the display luminance for each channel. The widely used GOG (gain-offset-gamma) model was proposed by Berns et al. (Berns, 1996). The model characterises the power-law relationship between the input digital signals and the output display luminance of the CRT (cathode ray tube) displays. Many LCD (liquid crystal display) manufacturers make the LCDs a power-law response mimicking the CRT displays thus the GOG models can also be used for LCD (Balasubramanian, 2017). The equation of the GOG model is given below:

$$R = (a_r \times d_r/(2^N - 1) + b_r)^{\gamma_r}$$
$$G = (a_g \times d_g/(2^N - 1) + b_g)^{\gamma_g}$$
$$B = (a_b \times d_b/(2^N - 1) + b_b)^{\gamma_b}$$

Equation 2.11

Where R, G, and B are linearised digital values ranging from 0 to 1; $d_r, d_g,$ and d_b are the input digital signals of R, G, and B channels; for each channel, a, b, γ are the coefficients of gain, offset, and gamma, and the sum of the gain and offset are constrained to 1. Neutral colours are recommended to determine the three coefficients (Roy S. Berns et al., 1993). Measurements of a set of neutral samples are required as the training

data and method of optimisation is needed to obtain the coefficients of gain, offset, and gamma for each channel with the minimum colour difference of the training data.

Other model also used to characterise the non-linear relationship include the PLCC (piecewise linear interpolation assuming constant chromaticity coordinate), which uses separate look-up-tables for each channel and assumes a linear relationship between points of the look-up-table (Post and Calhoun, 1989). Usually, the PLCC model requires more training data than the GOG model.

The second stage of display colour characterisation is a linear transformation between the linearised values and CIE XYZ tristimulus values, and the equation is given below:

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{r,max} & X_{g,max} & X_{b,max} \\ Y_{r,max} & Y_{g,max} & Y_{b,max} \\ Z_{r,max} & Z_{g,max} & Z_{b,max} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}
\qquad \textbf{Equation 2.12}
$$

Where R, G, and B are linearised digital values; X, Y, and Z are corresponding tristimulus values; $X_{r,max}$, $Y_{r,max}$, $Z_{r,max}$ are the tristimulus values when the red channel is at the maximum intensity, $X_{g,max}$, $Y_{g,max}$, $Z_{g,max}$ are the tristimulus values when the green channel is at the maximum intensity, $X_{b,max}$, $Y_{b,max}$, $Z_{b,max}$ are the tristimulus values when the blue channel is at the maximum intensity. This 3 x3 matrix can be obtained by three measurements of the pure red, pure green, and pure blue samples (Roy S. Berns et al., 1993).

2.6.3 Spatial independence and channel independence

When adopting the display characterisation models and applying the linear transformation, two underlying assumptions are made (Roy S Berns et al., 1993). The first important assumption is that the luminance and chromaticity of the colour displayed on one area of the screen are not affected by the colour on another area of the screen (spatial independence). To evaluate spatial independence, the white colour is usually measured against a black background and a white background, respectively, and the two measurements were compared in terms of the colour difference (Figure 2.19).
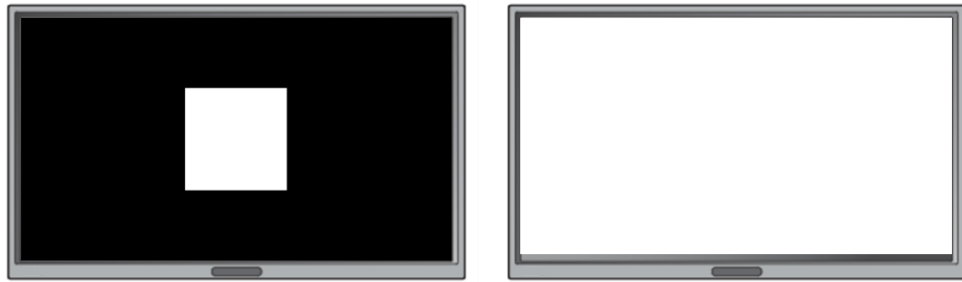
Figure 2.19 Evaluation of the spatial independence of display.

The second important assumption is that the output of each channel is independent (channel independence) so the output of the display is expected to be the sum of the tristimulus values of the three channels based on Grassmann's laws of additive mixing. The influence of the channel independence is tested by measuring the tristimulus values of the pure red, green, and blue colours (Figure 2.20), and comparing the sum of these tristimulus values with the tristimulus values of the pure white colour in terms of the colour difference.

Figure 2.20 Evaluation of the channel independence of display.

2.7 Psychophysics

Psychophysics is a discipline of the scientific study of the relationships between physical measurements of stimuli and the perceptions or sensations that those stimuli evoke (Gescheider, 2013). Psychophysical methods serve as tools to derive quantitative measures of all dimensions of human perceptions which are often considered subjective. It is the foundation of CIE colorimetry (Hunt and Pointer, 2011). As the present study deals with the human colour perception of facial skin, all the experimental work is associated with psychophysics. Specifically, visual psychophysics is used as the basis of the experimental work to deal with human colour perception.

There are two classes of visual psychophysical experiments: scaling experiments and matching or threshold experiments. The former is used to generate the relationship between physical measurements of stimuli and perceptual magnitudes. For example, scaling experiments could be used in

the evaluation of image quality. The latter is designed to measure the perceptual equality or the visual sensitivity of stimuli. Matching techniques are widely used in the study of colour appearance to determine whether two colour stimuli have a perceptually equal appearance across different conditions. It is critical to select the appropriate research method according to the specific situation. This section introduces the psychophysical scales and the two classes of methods.

2.7.1 Psychophysical scales

Four major types of scales are normally used in psychophysical experiments to describe the relationship between physical stimuli and the associated perceptions or sensations (Stevens, 1946).

Nominal scale

The nominal scales use numerals as labels or type numbers. They are categories without ordering or direction. For example, the numbering of players or classes, etc.

Ordinal scale

The ordinal scales are used to rank ordered categories. For example, the size of clothes, the classes of leather, etc.

Interval scale

The interval scales specify the differences between measurements but without a true zero point. For example, the Celsius temperature scale, the Likert scale for rating, etc.

Ratio scale

The ratio scales are interval scales with a true zero point. They are used in measurements of the ratio between a magnitude of a continuous quantity and a unit of measurement of the same kind. Most measurements in physical science and engineering use ratio scales. For example, the mass, the length, the duration, etc.

2.7.2 Scaling techniques

Different scaling techniques are used for the psychophysical study to quantitively measure the observers' responses. Three scaling techniques are introduced in this section, categorical judgement, pair comparison, and magnitude estimation.

2.7.2.1 Categorical judgement

Categorical judgement is an effective scaling method. Observers are asked to respond with their opinion of the stimuli by an equal–interval scale of categories, e.g. a seven-point Likert scale. Torgerson's Law of Categorical Judgment is applied to this method (Torgerson, 1958). Unlike the paired comparison, the method is useful particularly when the number of stimuli is large. Previous studies on skin colour and facial preference judgements have largely used two methods, categorical judgement and pair comparison (see Section 2.4, as summarised in Table 2.1 and Table 2.2). A few also had observers adjust the colour of the stimuli. The number of references using each method is listed in Table 2.4 below. Pair comparison could be useful for studies using manipulated facial images as other variables are controlled and only subtle colour changes exist between sample pairs. While studies using the uncontrolled stimuli of real facial images have mainly chosen the method of categorical judgement and the 7-point. In the present study, the 7-point Likert scale was also used for the preference evaluation based on facial skin colour (see Experiment 1 and Experiment 2).

Table 2.4 Psychophysical techniques used in the literature on skin colour and facial preference judgements.

Methods	Specific techniques	References amount
Categorical judgement	Preference rating using a 7-point Likert scale	13
	Preference rating using a 9-point Likert scale	5
	Other techniques, e.g. 10-point rating scale, 0-100 scale, questionnaire, etc.	3
Pair comparison	Forced choice between pairs of stimuli, e.g. participants were told to choose the face they thought was more preferred	8
Adjusting colour	change the colour to make the face as attractive/healthy as possible	4

2.7.2.2 Pair comparison

As mentioned above, paired comparison can be used in the experiment with a relatively small number of stimuli. The observers will view all the possible pairwise combinations of stimuli and are usually asked to choose one stimulus after evaluation. The proportion of times a particular stimulus is chosen is calculated and recorded during the experiment. Thurstone's Law

of Comparative Judgement is applied to this method (Thurstone, 1927). Examples of pair comparison can be found in the literature reviewed in Section 2.4 as summarised in the last column of Table 2.1 and Table 2.2.

2.7.2.3 Magnitude estimation

Magnitude estimation is a technique standardly applied in psychophysics to measure judgments of sensory stimuli (Teghtsoonian et al., 1975). In magnitude estimation, observers judge the intensity of a stimulus in comparison to a baseline stimulus or reference stimulus. This method can directly generate the ratio scale and is highly replicable both within and across observers, thus has been widely adopted to study colour appearance. In fact, magnitude estimation and matching are the two suggested psychophysical techniques for colour appearance modelling (Fairchild, 1995; Fairchild, 1997). To study the colour appearance, observers are asked to estimate the scale values of the colour appearance attributes, e.g. lightness, colourfulness, hue etc. (see Section 2.2.4 for perceptual attributes of colour) based on a reference stimulus. Using this method, the perceptual scale values can be directly obtained in the context of various parameter settings. This method is used for accumulating some most important colour appearance datasets, e.g. LUTCHI data (Luo et al., 1991) as well as developing and testing the colour appearance models (Luo et al., 1991; Fairchild, 1995; Luo and Hunt, 1998; Kuo, 2007; Pointer et al., 2008). The advantage of this method is that the training is easy, and observers can easily scale the perceptual attributes. The disadvantage is the variation of observers could be larger than the matching techniques. Hence, the coefficient of variation (CV) has been suggested as a statistical measure to compare the observer variations in data analysis (Fairchild, 1995). The calculation of the CV value will be introduced in Section 2.8.2 for measures of observer variation. In the present study, the magnitude estimation is used to scale three perceptual attributes of facial colour appearance, whiteness, redness, and yellowness based on the reference stimuli, and the CV value is used to assess the observer variation (see Experiment 4).

2.7.3 Matching techniques

Colour matching techniques are the other widely used method in studies on colour appearance and there are two major types of matching techniques, asymmetric matching and memory matching.

2.7.3.1 Asymmetric matching

Asymmetric colour matching is often used in the study of colour appearance assessment to determine when two stimuli are not perceptually different (Fairchild, 2013). For example, the CIE colour matching functions are derived from colour matching experiments, providing the basis of CIE colourimetry. When judging adjacent colour samples, the human visual system is good at detecting whether they are equal or not. The observers produce a colour match while viewing both stimuli and then the physical properties of the match will be used to study the perception of the human visual system. All the studies reviewed in Section 2.5.2 have used the technique of asymmetric matching to study the overall colour appearance of multi-coloured objects (Kuriki, 2004; Sunaga and Yamashita, 2007; Giesel and Gegenfurtner, 2010; Kimura, 2018; Virtanen et al., 2020). In the present study, this method is used to assess the overall colour appearance of human faces (see Experiment 3).

2.7.3.2 Memory matching

Memory colour matching is another type of matching technique used in colour appearance research. The observers are asked to match the colour of a familiar object under adaptive conditions until it matches the observers' memorized colour of this object. Hasantash et al.'s study gives an example of memory colour matching (Hasantash et al., 2019).

2.8 Data analysis techniques

The statistical methods and modelling techniques used for data analysis are introduced in this section. The data analysis in the present study is conducted using R (RDC, 2010), which is a programming language for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing.

2.8.1 Statistical measures and tests

Several statistical measures and tests described in this section served as the bedrock of data analysis. The book, *Introduction to statistics in psychology* by Howitt and Cramer (Howitt and Cramer, 2007), is used as a general reference for this section.

The mean

The arithmetic mean is adopted in this study, which is the everyday concept of average. It is calculated by summing all of the data in a distribution and

then dividing by the number of the dataset. The statistical notation can be expressed in the equation below:

$$\hat{X} = \frac{\sum X}{N}$$

Equation 2.13

Where $\hat{X}$ is the statistical symbol for the arithmetic mean of a set of data; $\sum X$ means add up all of the data X, N is the number of the data.

The median

The median is the middle data of a set if the dataset is organised from the smallest to the largest.

The variability

The concept of variability is essential in statistics. The statistical term, variance, is the basic measure of variability.

$$variance = \frac{\sum(X - \hat{X})^2}{N}$$

Equation 2.14

The standard deviation

The standard deviation measures the dispersion of a set of data relative to its mean and is calculated as the square root of the variance.

$$\sigma = \sqrt{\frac{\sum(X - \hat{X})^2}{N}}$$

Equation 2.15

Data standardisation

Based on the mean and standard deviation, data standardisation can be achieved by calculating the Z-scores, also referred to as standardised scores or standard scores. Z-score standardisation refers to the process of standardising every value in a dataset such that the mean of all of the values is 0 and the standard deviation is 1, as expressed in the equation below. It allows all the data to be expressed consistently in a standard form so that the variables with many different units of measurement can be compared.

$$standardised\ score = (X - \hat{X})/\sigma$$

Equation 2.16

The benefit of standardisation is the clear outliner can be transformed so that it's no longer a massive outliner. Otherwise, the outliner will largely influence the model fit especially for some types of machine learning models. thus, in the present study, the dataset including various facial colour characteristics is standardised prior to the modelling analysis (see Chapter 5 and Chapter 6).

Person correlation coefficient

With two or more sets of data (variables), the relationships between them become important. The relationships can be visualised by the scatterplots or quantified by correlation coefficients. Correlation coefficients are the numerical indices measuring the relationship between two sets of data. The Person correlation coefficient, *r*, is primarily used for score variables, and is determined by the equation below.

$$r = \frac{\sum(X - \hat{X})(Y - \hat{Y})}{\sqrt{\sum(X - \hat{X})^2}\sqrt{\sum(Y - \hat{Y})^2}}$$ **Equation 2.17**

The coefficients can be either positive (the two sets of data increase together) or negative (one set of data increases as the other decreases). The absolute values of the coefficients range from 0 (no relationships) to 1 (perfect relationships), indicating how close the relationships are. In the present study, Person correlation is used in the data analysis of different experiments. The Person correlation coefficients are reported along with the statistical significance of the correlation coefficients. The correlation tests are conducted using the *cor()* function of the *stats* package in R.

The t-test

The *t*-test is mainly used to assess the difference between two sets of data which are collected under two separate conditions but from a single sample of participants. It is also appropriate to apply the *t*-test when the two sets of scores are correlated with each other as when matching is used, which is the situation in Experiment 3 (colour appearance matching) of the present study. A one-sample *t*-test is a simple type of *t*-test used to determine whether or not the mean of a population is equal to some value. It is used in the analysis of Experiment 3 (see Chapter 7). The one-sample *t*-test was done using the *t.test()* function in the *stats* R package.

The linear mixed-effect analysis

The linear mixed-effect analysis is an extension of the simple linear analysis. The term mixed refers to its capability to allow both the fixed effects and random effects to be tested in the same analysis. The mixed analysis provides a general, flexible approach to correlated data or hierarchical structures (observations with subgroups). The fixed effects parameters tell how population means differ between any set of treatments, while the random effect parameters represent the general variability among subjects or other units. In the present study, the linear mixed-effect analysis is used in Experiment 1 to test the fixed effects of L*, a*, and b*, image ethnicity and

observer ethnicity, and meanwhile include the random intercepts for both images and observers. The linear mixed-effect analysis was implemented by the *lmer()* function in the *lme4* R package (Bates et al., 2015). As image ethnicity and observer ethnicity are categorical predictors, the deviation coding was used to convert both image ethnicity and observer ethnicity into deviation-coded factors. The detail of the analysis can be found in Chapter 4.

2.8.2 Measures of observer variation

In visual psychophysical experiments, the observer variation, including the intra-observer and the inter-observer variation, is critical and needs to be initially considered to be aware of the typical error involved in the experiments. The intra-observer variation is the variation in repeated measurements by the same observer, also referred to as observer repeatability. The inter-observer variation is the variation of measurements between different observers, also referred to as observer accuracy. According to the psychophysical techniques used, different types of data are collected in the four experiments of the present study. Thus, different measures are used to evaluate the observer variation, which is also for the purpose of comparing the variation with the existing literature. Three methods are introduced in this section to quantify the observer variation in the present study. The observer variation of each experiment will be calculated in the next chapter.

The Cronbach's alpha coefficients

Cronbach's alpha is a measure of internal consistency or reliability, which shows how closely related a set of items is as a group (Cronbach, 1951). It used to be a measure of the reliability between scales, but it has been widely adopted as a measure of consistency between ratings from different observers (inter-observer variability) in a large number of studies on skin colour facial preference judgements (as reviewed in Section 2.4). It is expressed in the following equation:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Equation 2.18

Where N is the number of items; $\bar{c}$ is the average inter-item covariance among the items and $\bar{v}$ is the average variance. The values of the Cronbach's alpha coefficients range from 0 to 1 where a higher number indicates a greater consistency or smaller variation. In the present study, Experiment 1 and Experiment 2 collect the ratings of preference judgements

and thus use the Cronbach's alpha coefficients as the measure of observer variation (see Section 3.3.1.2 and Section 3.3.2.2).

The mean colour difference from the mean (MCDM)

The mean colour difference from the mean (MCDM) is a measure commonly used to describe colour variations for a set of data points in CIELAB space using the following equation (Nadal et al., 2011; Berns, 2019),

$$MCDM = \frac{\sum_{i=1}^{N}[(L_i^* - \bar{L}^*)^2 + (a_i^* - \overline{a^*})^2 + (b_i^* - \overline{b^*})^2]^{1/2}}{N} \qquad \textbf{Equation 2.19}$$

where L_i^*, a_i^*, and b_i^* are the CIELAB coordinates for the ith measurement, $\bar{L}^*$, $\overline{a^*}$, and $\overline{b^*}$ are the average CIELAB coordinates and N is the number of measurement. Larger values of MCDM indicate larger colour differences or larger variations. In the present study, the datasets collected in Experiment 3 (colour appearance matching) are the matched colours all specified in CIELAB coordinates. Thus, the MCDM is adopted in experiment 3 to assess the observer variability including the inter- and intra- observer variability (see Section 3.4.1.2). Besides, the MCDM is also used in this study to evaluate skin colour variations of target facial areas (see Section 5.2.1).

The coefficient of variation CV

The coefficient of variation CV is a statistical measure to represent the agreement between two sets of data and is expressed in the equation below.

$$CV = 100 \left[\sum (x_i - k * y_i)^2/n\right]^{1/2}/\bar{y} \qquad \textbf{Equation 2.20}$$

Where n is the number of samples in x and y datasets; $\bar{y}$ is the mean value of the y dataset; k is a scaling factor and is equal to 1 when the two datasets have the same unit. To determine the inter-observer variation, x_i and y_i are the individual results and the mean results over all observers; for the intra-observer variation, x_i and y_i are the first and second judgement data, respectively. The CV value can be thought of as the relative percentage error, For a perfect agreement, the CV value should equal 0. Larger values indicate worse agreement or larger variations. The CV is recommended by CIE TC 1-34 to assess the observer variation for the technique of magnitude estimation in colour appearance scaling (Fairchild, 1995; Fairchild, 1997). It has been widely implemented in colour appearance research (Luo et al., 1991; Fairchild, 1995; Luo and Hunt, 1998; Kuo, 2007; Pointer et al., 2008). In the present study, the magnitude estimation is used in Experiment 4 to scale the perceptual attributes of facial colour appearance, and the CV value

is used to assess the observer variation including the inter- and intra-observer variability (see Section 3.4.2.2).

2.8.3 Modelling techniques

Based on the basic statistical measures and analysis, it is helpful to identify the relationships between the data and make further predictions by applying modelling techniques. Different models are applied to the present study to fit the different types of the data and purpose of modelling. For example, the relationship between one independent variable and one dependent variable is estimated by the simple linear regression model (e.g. predict the perceived face colour from the average face colour, see Chapter 7). A multiple regression model is used when several explanatory variables are involved (e.g. quantify the perceptual attributes of facial colour appearance from the L^*, a^*, and b^* values, see Chapter 8). However, these models could be inaccurate to handle a larger number of explanatory variables and the complex relationship between them (high dimensional data) (Filzmoser and Nordhausen, 2021), e.g. to model the relationship between a large dataset of facial colour characteristics and the response variable of facial preference. Hence, some other multivariate statistical techniques are considered and their effectiveness is discussed (see Chapter 5 and Chapter 6). In this section, different modelling methods that are used in the present study are introduced.

Ordinary Least Squares Regression (OLS)

OLS has been widely used as a standard method to predict a response variable from one or several explanatory (predictor) variables. The parameters in OLS models were estimated by minimizing the residual sum of squares. OLS with only one predictor is called simple linear regression, while with two or more predictors is multiple linear regression. Multiple linear regression (MLR) is one of the most widely used methods for forecasting.

Running an OLS model with too many variables, especially irrelevant ones, will lead to a needlessly complex model. One way to simplify the model and reduce the multi-collinearity is to perform subset selection. Stepwise regression is the multiple linear regression with a process of subset selection. The method adds the most relevant variable or removes the least relevant predictors from the model in each step so that only the most important ones that are capable of predicting the variation in the response variable are left in the final model. Two stepwise methods that are mostly used are forward and backward methods. Stepwise Regression (Forward

Steps) starts with only one intercept and then adds variables based on certain criteria (e.g. AIC, correlation significance, etc.) in a stepwise manner until no variable can be added (Yamashita et al., 2007). The forward method is more suitable when the number of variables is larger than the sample size. Stepwise Regression (Backward Steps) starts with a full model and then removes one variable in each step based on certain criteria until the model cannot remove any more variables. As the backward elimination method has the advantage of considering the effects of all the colour variables simultaneously at the start point, it generally has better performance than the forward methods when the number of variables is not too large.

The OLS method has been most widely used for forecasting in many research fields such as the environment (Abdul-Wahab et al., 2005; Çamdevýren et al., 2005; Pires et al., 2008; Mountains, 2013; Gomes et al., 2014), economics (Chan and Park, 2005), and psychology (Ansiau et al., 2005). Despite its success in many applications, yet the regression methods can face difficulties when the number of explanatory variables is too large and there are collinearities between them (McAdams et al., 2000). In such cases, different variable selection methods in MLR analysis may lead to different models and make it difficult to identify the most important contributor in the model.

Dimension reduction regression

The idea of Principal Component Analysis (PCA) is to linearly transform the given features inside the feature space into a specific principal component space which maximizes the variance and minimises the number of features utilised to explain such variance. Based on the PCA, principal component regression and partial least squares regression are the common approaches that adopt the pre-processing to remove the dependence of explanatory variables, and thus avoid overfitting. Principal component regression (PCR) overcomes the problem of multi-collinearity by projecting the original predictors into an uncorrelated subspace of principal components and then using OLS to fit the linear regression model using the principal components as predictors (Pires et al., 2008; Abdi and Williams, 2010). In PCR, the dependent variable is not used to identify each principal component direction, thus it is not guaranteed that the principal components are related to the response variable. Partial least squares regression (PLSR) is the supervised version of PCR as it projects the original predictors into a subspace of latent components that maximize the covariance between the response variable and the predictor variables (Geladi and Kowalski, 1986).

Regularisation regression

Regularisation of penalisation is a relatively new and entirely different approach to deal with multicollinearity. Techniques of regularisation or penalisation are used to constrain the coefficients of a model and reduce the variance of the parameter estimators. The most well-known Regularisation regression methods are ridge, lasso, and the elastic net. The Ridge Regression (RR) is essentially a linear regression model but with added constraints to the coefficients (Hoerl and Kennard, 1970; McDonald, 2009). All coefficients shrink towards 0 by a tuning parameter and the model still maintains the structure of the original OLS model. In contrast, Least Absolute Shrinkage and Selection Operator Regression (LASSO) is another form of regularisation regression which penalises high coefficients and forces coefficients to be 0 (whereas RR make coefficients approach 0) (Tibshirani, 1996). LASSO works well for feature selection as it takes out unnecessary variables from the model and makes the model simpler and more interpretable than RR. The Elastic Net Regression (EN) combines shrinkage from RR and LASSO and balances the two algorithms by weighting the two effects. EN is a linear regression which shrinks predictors to reduce overfitting through Regularisation and meanwhile performs variable selection by setting the coefficients of uninformative parameters to zero (Hastie et al., 2001; Zou and Hastie, 2005). The models have two hyperparameters, α and λ, which could be tuned to optimize the model fit. α controls the degree to which the model shrinks coefficients and λ determines how aggressively coefficients are set to zero. The implementation of this approach can be found in previous studies (Jaeger and Jones, 2022; Jaeger and Meral, 2022).

Cross-validation is often used in those regularisation regressions to determine the tuning parameter with maximised model fit and to separate the training and testing of a dataset for model performance evaluation. In K-fold cross-validation, the dataset is randomly divided into K folds, one of the folds is the holdout set and the remaining K-1 folds are to fit the model. The model performance is evaluated, usually by RMSE, using the fold that is held out. The process can be repeated using a different set each time as the holdout set, and the overall model performance is the average over all splits. The advantage of using cross-validation is that the dataset is split and the model performance is tested using an independent testing dataset. With repeats of different splits, it is more likely to obtain an unbiased estimate of the model performance.

2.8.4 Model selection and performance evaluation

In order to compare the overall model performance between different multivariate modelling techniques, the coefficients of determination and the root mean squared error are introduced.

The coefficients of determination R^2

The coefficients of determination, R^2, measures how well the predictions of a statistical model fit a set of observations. R^2 is calculated as the square of the Person correlation coefficients between the observed values and the model-fitted values. In most cases, R^2 ranges from 0 to 1, and a higher value indicates a better model fit. As more explanatory variables are added to the model, the R^2 automatically increases. The adjusted R^2 is often used to correct the goodness-of-fit measure for linear models, and it can be calculated by the following equation.

$$adj.\,R^2 = 1 - (1 - R^2)\frac{n-1}{n-p}$$ **Equation 2.21**

Where p is the total number of explanatory variables in the model and n is the sample size. The adjusted R^2 can be largely influenced by the number of explanatory variables.

Root Mean Squared Error (RMSE)

The Root Mean Square Error (RMSE) of prediction is the widely used criterion to represent the prediction error using the same units as the original rating scale. It is defined as the equation below:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$ **Equation 2.22**

Where $\hat{y}_i$ are the predicted values, and y_i are the observed values. A smaller value of RMSE indicates better goodness of fit of the model. Compared to statistics such as the R^2, RMSE is not inflated by the number of predictors.

2.9 Summary

This chapter gave a review of the relevant literature to the present study. The key information is summarised below.

- The knowledge of human colour perception was reviewed. Colour is more than a physical stimulus. The perceptual aspects and the

psychological cognition of the human visual system largely influence the perception and preference judgements of skin colour.

- The CIE colorimetry provides the numeric system for colour specification. CIELAB uniform colour space was adopted in the present study for skin colour specification, facial appearance investigation, data analysis and model development.

- Facial colour preferences were widely explored using image-based methods. Various facial colour characteristics affect preference judgements. The limitations of the existing studies include: 1) results were controversial due to the different methods used while studies using realistic skin models lacked; 2) different colour characteristics were examined in isolation, it was unknown how they took together would influence preference judgements and what their relative importance was; 3) the cultural differences were not satisfactorily understood.

- Research on facial colour preference showed the importance of facial colour appearance. Although CIE colorimetry provides the objective tool for colour measurement and quantification, colour perception is subjective in nature. The existing studies suggested: 1) the overall facial colour appearance was not clear; 2) the perception of facial whiteness, redness, yellowness, and their relationships with the colorimetric values were not precisely examined; 3) the perceptual difference between face and nonface object existed but was not fully understood.

- Display colour characterisation is essential to achieve accurate and consistent colour reproduction of the visual stimuli on display. The characterisation process was implemented in the displays used for appearance assessment and preference evaluation in the present study.

- Psychophysics provides fundamental tools for this study to derive quantitative measures of subjective colour perceptions related to facial skin. In this study, the technique of categorical judgement was adopted to measure the preference evaluation; the asymmetric colour matching was used to assess the overall colour appearance of human faces; magnitude estimation was implemented to scale three perceptual attributes of facial colour appearance, whiteness, redness, and yellowness.

- The statistical methods and modelling techniques used in this study were introduced. Different methods and techniques were chosen for

different data analyses according to the types and characteristics of datasets collected as well as the purpose of the experiments.

Chapter 3 Experiments

3.1 Overview

In this chapter, psychophysical experiments for assessing facial colour preference (Experiments 1, 2) and facial colour appearance (Experiments 3, 4) were described. Figure 3.1 shows an experiments overview. Experiment 1 and 2 aimed to investigate the relationship between skin colour and facial preference judgements using categorical judgement methods. Experiment 1 was a cross-cultural study conducted in the UK, and Experiment 2 was a validation experiment conducted in China. Experiments 3 and 4 aimed to investigate the human perception of facial colour appearance. Experiment 3 studied the overall facial colour appearance using the technique of asymmetry colour matching. Experiment 4 studied perceptual facial whiteness, redness, and yellowness using the technique of magnitude estimation. The experiments for this study were carefully prepared to achieve a reliable and consistent result. The details of experimental preparations were reported, including display (Section 3.2.1) and image (Section 3.2.2). Details of each of the four experiments were then given in Section 3.3.1, Section 3.3.2, Section 3.4.1, and Section 3.4.2, respectively. Within each section, the experiment design and procedure were described, followed by the observers and the evaluation of observer variations. The results of experiments will be presented and discussed in the following Chapters 4-8.
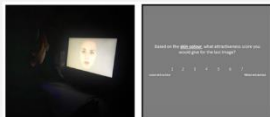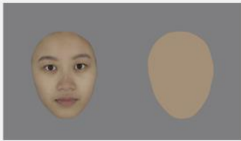
	Task	Display	Stimulus	Observer	Data analysis
Experiment 1	Preference evaluation (attractiveness, healthiness, age)	BenQ	80 facial images 40 CA + 40 CN	44 Observers 22 CA + 22 CN	Chapter 4 Chapter 5
Experiment 2	Preference evaluation (attractiveness)	iPad	100 facial images 40 CN + 60 CN (new)	51 Observers 51 CN	Chapter 6
Experiment 3	Appearance matching	BenQ	80 facial images 40 CA + 40 CN	44 Observers 21 CA + 23 CN	Chapter 7
Experiment 4	Appearance scaling	BenQ	80 facial images 80 uniform patches 40 CA + 40 CN	43 Observers 23 CA + 20 CN	Chapter 8

Figure 3.1 Experiments overview (CA = Caucasian, CN = Chinese).

3.2 Experimental preparation

3.2.1 Display

Experiments 1, 3, and 4 were conducted at Leeds using a 24.1-inch BenQ colour professional display (51.84 cm x 32.40 cm, LCD backlit, adobe RGB colour space). Experiment 2 was conducted in Shanghai, China and used a 12.9-inch iPad pro display (28.06 cm x 21.49 cm, LED backlit, P3 colour space) due to the difficulty of moving the BenQ display from Leeds to Shanghai. During the experiments, the BenQ display was placed horizontally and the iPad display was placed vertically so that the facial images were presented in a similar size on both displays. Before the experiments, the colour characteristics of both displays were measured and characterisation was done to precisely control the colour and consistently and accurately reproduce the appearance of human faces on both displays during the experiments.

3.2.1.1 Colour measurement

All the measurements were carried out in a dark room. Before the radiometric measurements, the displays were placed where they would be used and then turned on for 30min to warm up. The Konica Minolta CS-2000 spectroradiometer (1° field of view) was used to measure the colour on display (BenQ and iPad) in term of the CIE XYZ tristimulus values in unit of cd/m^2. The measured results were used to evaluate the display characteristics and develop colour characterisation model. Figure 3.2 shows the measuring geometry, the spectroradiometer was placed at a distance of 0.6 m away from the display, and all the measurements were performed at the central point of the display. All images for characterisation were made by Microsoft PowerPoint software and displayed in full-screen mode. The absolute tristimulus values for 2° observers were obtained.
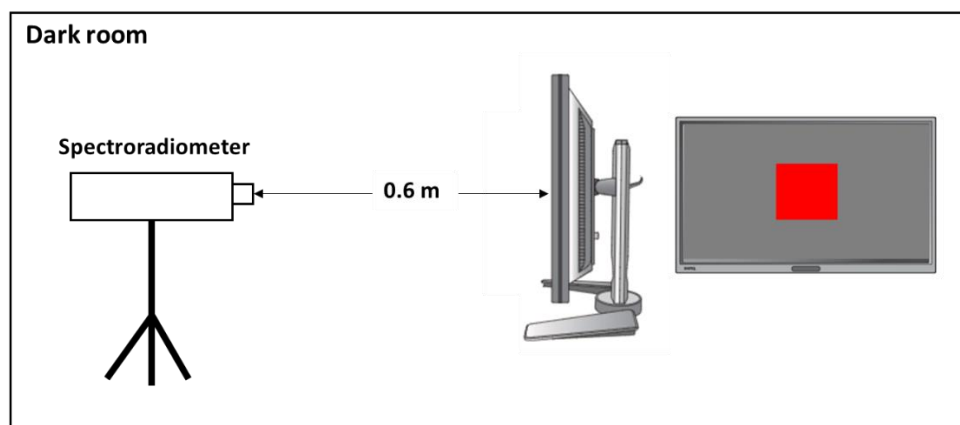
Figure 3.2 Colour measurement geometry for display characterisation.

3.2.1.2 Spatial independence and channel independence

The spatial independence of the display was first assessed by measuring the CIELAB colour difference between a white patch with a black surround, in the centre of the display, and the same white patch with a white surround. The results were 0.19 ΔE_{ab}^* for the BenQ display, and 0.84 ΔE_{ab}^* for the iPad display, which indicated a good spatial independence for both displays. The channel independence was then assessed by the colour difference between a full-field white and the prediction of the full-field white based on the sum of the tristimulus values of the full-field pure red, green, and blue. The colour difference was 1.41 ΔE_{ab}^* for the BenQ display, and 2.01 ΔE_{ab}^* for the iPad pro display. Both results indicated relatively good channel independence.

3.2.1.3 Display characterisation

The GOG (gain-offset-gamma) model was employed to carry out the display colour characterise for both displays (Burns and Berns, 1996; Day et al., 2004). The pure red, green, blue and 18 greyscale patches (equal dr, dg, db values of 0, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 195, 210, 225, 240 and 255) seen against a neutral grey background, were used as the training dataset for building the GOG model. The mean colour difference between the instrument measurement and the model prediction of the 18 training samples was calculated to evaluate the accuracy of the display model. The optimisation function *solver* in Microsoft Excel was then used to solve the coefficients of gain, offset and gamma for each channel by minimising the average colour difference ΔE_{ab}^* of the 18 training samples. After the optimisation, the coefficients of gain, offset and gamma for each channel of the two displays were obtained as listed in Table 3.1. The BenQ display model generated a mean colour difference of 0.48 ΔE_{ab}^* within the training dataset, and the iPad pro display model generated a mean colour difference of 0.08 ΔE_{ab}^*. Figure 3.3 and Figure 3.4 show the relationships between the digital input and the luminance output and the CIE tristimulus values of the two displays.

Table 3.1 Gain, offset and gamma for each channel of the BenQ display and the iPad display.

	Gain	Offset	Gamma
BenQ display			
Red	0.9964	0.0001	2.2605
Green	1.0125	-0.0051	2.2447
Blue	1.0129	-0.0096	2.2442
iPad display			
Red	0.9949	-0.0039	2.1731
Green	1.0070	-0.0030	2.1826
Blue	1.0071	-0.0033	2.1797
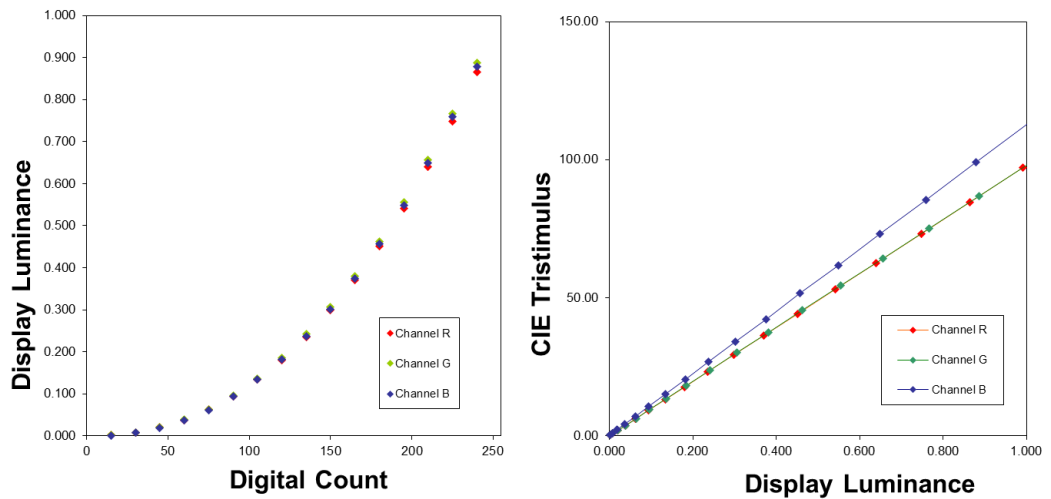
Figure 3.3 The relationships between the digital input and the luminance output and the CIE tristimulus values of the BenQ display.
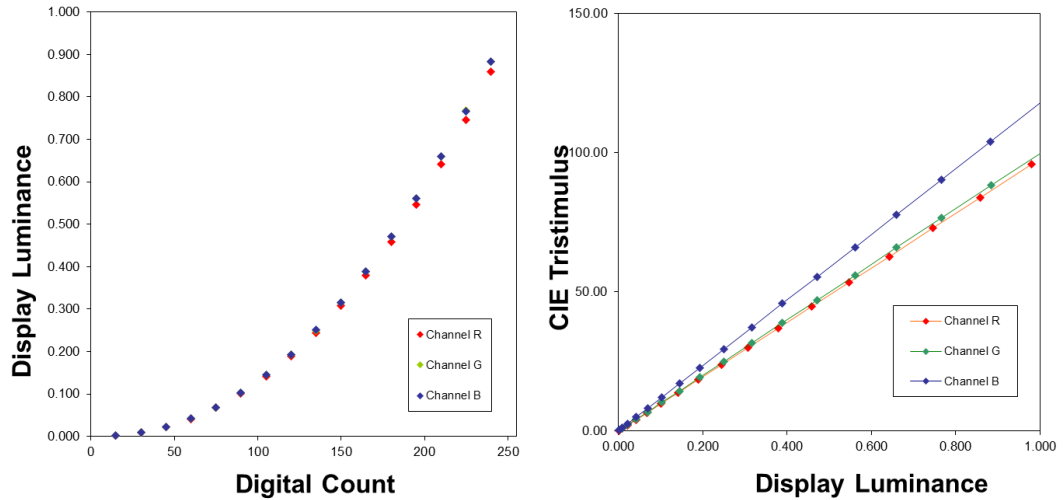
Figure 3.4 The relationships between the digital input and the luminance output and the CIE tristimulus values of the iPad display.

To evaluate the model performance on skin colour reproduction, twenty random-selected skin colour patches, again seen against a neutral grey background, were used as the testing dataset. The BenQ display model generated a mean colour difference of 0.80 ΔE_{ab}^* averaged from the 20 skin colour samples as testing colours; The iPad pro display model generated a mean colour difference of 0.21 ΔE_{ab}^* for the testing skin colours. Because the GOG model gave a relatively good colour reproduction performance, other models for display colour characterisation were not considered in this study. The models were used to process the facial images by computing the RGB values that are required to display any colour specified in CIE colorimetric values (see Section 3.2.2.2).

3.2.2 Image

3.2.2.1 Image photography and selection

All facial images used in this study were selected from the Liverpool-Leeds Skin-colour Database (LLSD), which included both the image data and corresponding reflectance data of each pixel in the images. The facial image of each subject was obtained by photography in a VeriVide DigiEye® light booth, which provided a uniform matte mid-grey background and even, diffuse, fluorescent illumination that simulated CIE illuminant D65. There was no other lighting in the room where the photography took place. During data collection, the participant sat in the viewing cabinet and their target facial area was adjusted to fit within the camera image. A digital SLR camera (Nikon D7000), controlled by the DigiEye® system software, was used to capture each subject's face. The distance from the participant to the camera

was fixed at approximately 57.5 cm and the participant looked straight into the camera.

Eighty real facial images, including 40 Caucasian images and 40 Chinese images, all with neutral facial expressions, were selected from the LLSD database for all the experiments in this study. Apart from the same 40 Chinese images, Experiment 2 used additional 60 Chinese images as a validation dataset, and these images were also selected from the LLSD database.

3.2.2.2 Image processing

For each facial image, the spectral reflectance data of each pixel was first transformed to CIE XYZ values by Equation 2.1 using the CIE illuminant D65 as the light source, and the tristimulus values were then converted to display RGB values based on the display colour characterisation model. The BenQ display model was used to process the 80 facial images to be presented on the BenQ display in Experiment 1, 3, and 4; The iPad Pro display model was used to process the 100 Chinese images used in Experiment 2. After processing the colour, the hair, ears, and any visible clothing were then manually removed from each image using Photoshop CS6. The image was at last scaled to be in the centre of the screen with a mid-grey background (L^*, a^*, b^* = 50, 0, 0 in CIELAB colour space). Figure 3.1 shows an example of a Chinese real facial image and a Caucasian real facial image used in this study.

Figure 3.5 An example of a Chinese real facial image (left) and a Caucasian real facial image (right).

3.3 Experiments on facial colour preference

3.3.1 Experiment 1

3.3.1.1 Study design

Experiment 1 was a psychophysical experiment of facial preference evaluation using the categorical judgement methods. It was conducted in Leeds, UK. The 80 facial images of real human faces, including 40 Caucasian images and 40 Chinese images, were used in this experiment.

The experiment was conducted in a dark room and using a self-compiled MATLAB program to display facial images on the BenQ display. There was a 3-min dark adaptation before the formal experiment. After that the observers were given the experimental instructions. The experiment was divided into three separate sessions. In each session, the observer viewed 80 facial images presented in random order and rated each skin colour of each image concerning one of the three attributes, perceived attractiveness, perceived healthiness, and perceived visual age.

Figure 3.6 shows the experimental scene and one of the questions. Observers had eight seconds to view each facial image, then the question page would automatically show on screen. The following question was asked after the observation of each image, "Based on the **skin colour**, what attractiveness score (or healthiness score or the estimated age, depend on different sessions) you would give for the last image?" The observer then made a judgement of the facial skin colour without a time limit. The ratings were recorded manually by the experimenter. After rating one image and clicking the mouse, the next facial image would appear on display automatically.
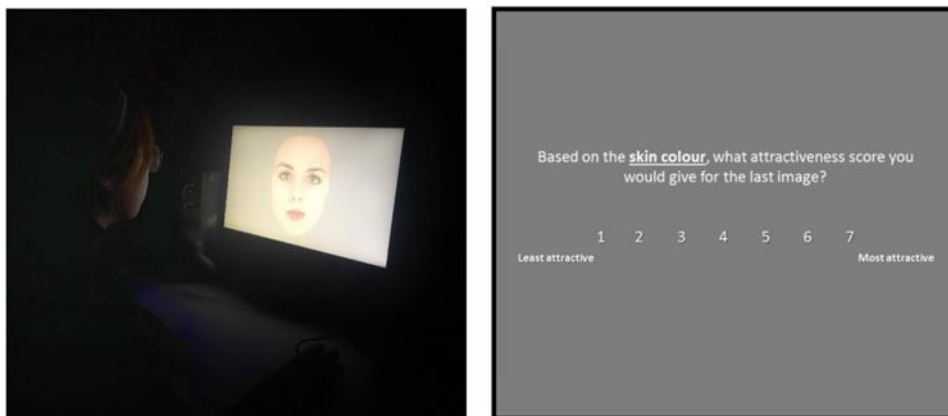
Figure 3.6 Observer evaluating the facial image on display (left) and one of the questions shown on display (right).

Using a categorical judgment method, the perceived facial attractiveness and healthiness were rated on a 7-point Likert-type scale where 1 represented 'least attractiveness' / 'healthiness' and 7 represented 'best attractiveness' / 'healthiness'. The visual age was rated on a single-year step scale from 1 to 99 years.

3.3.1.2 Observers and observer variations

The experiment used 44 observers with normal colour vision, including 22 Caucasians (13 males; mean age $\pm$ SD = 24.27 $\pm$ 5.30) and 22 Chinese (7 males; mean age $\pm$ SD = 26.05 $\pm$ 3.96), whom each evaluated the colour appearance of the 80 facial images using three subjective attributes. The Chinese observers were from mainland China, and at the time of the study, they were at Leeds University, UK as students or visiting scholars. On average, they spent about 1–3 years in the UK. All the observers were given instructions in English and confirmed their understanding before the experiments.

The Cronbach Alpha Coefficient was used to assess the inter-observer variability of the Caucasian and Chinese observers (Cronbach, 1951). As shown in Table 3.2, the values of Cronbach's alpha coefficient are all greater than 0.87, suggesting that there is high internal consistency in the judgements of attractiveness, healthiness, and age for both the Caucasian and Chinese groups of observers. For both sets of observers, inter-observer variability was a little higher when rating their own ethnic's images. Both sets of observers agreed more strongly on what is attractive, healthy and young in their own-ethnicity faces compared to other-ethnic faces.

Table 3.2 The Cronbach Alpha Coefficient for assessing the inter-observer variability of the Caucasian (CA) and Chinese (CN) observers (sample size).

	CA	CN	CA & CN
CA images			
Attractiveness	0.96 (22)	0.93 (22)	0.96 (44)
Healthiness	0.96 (22)	0.93 (22)	0.97 (44)
Age	0.90 (22)	0.91 (22)	0.95 (44)
CN images			
Attractiveness	0.95 (22)	0.96 (22)	0.97 (44)
Healthiness	0.96 (22)	0.96 (22)	0.98 (44)

| Age | 0.87 (22) | 0.92 (22) | 0.94 (44) |

3.3.2 Experiment 2

3.3.2.1 Study design

Experiment 2 was conducted in Shanghai, China to obtain another set of facial attractiveness evaluation data using a new set of Chinese facial images and a new panel of Chinese observers, which was used as a validation dataset for Experiment 1. One hundred Chinese facial images were used as the visual stimuli in Experiment 2, including the same 40 images as Experiment 1, which were used to test the observer consistency between Experiment 1 (Leeds) and experiment 2 (Shanghai), and another 60 Chinese facial images as the new testing materials.

The experiment was conducted in a dark room. The experimental procedure was the same as Experiment 1 but only facial attractiveness was assessed. The observers viewed each of the facial images presented in a random order and then made a judgement of perceived facial attractiveness based on the skin colour using a 7-point Likert-type scale where 1 represented 'least attractiveness' and 7 represented 'best attractiveness'. During the experiments, to present the facial images in a similar size to the BenQ display, the iPad display was placed vertically.

3.3.2.2 Observers and observer variations

Fifty-one Chinese observers (21 males, mean age ± SD = 24.45 ± 4.10) including both genders with normal colour vision participated in experiment 2, and at the time of the study, they were all native Chinese and were either students or researchers at Fudan University, China.

Cronbach's alpha was again calculated as a measure of agreement for attractiveness ratings across observers. The internal consistency was also high for ratings in Experiment 2: Cronbach's alpha = 0.98, 95% CI [0.98, 0.99]). Furthermore, the rating results of the same forty Chinese images between the two experiments were compared to assess the consistency between the two groups of Chinese observers and the two sets of rating scores were found to be significantly highly correlated (Person's correlation coefficient: $r(38) = 0.94$ [0.89, 0.97], $p < 0.001$). The high consistency of the preference judgement results between the two experiments showed that the short experience of living abroad would not affect the aesthetic preference of Chinese observers.

3.4 Experiments on facial colour appearance

3.4.1 Experiment 3

3.4.1.1 Study design

Experiment 3 was a colour appearance matching experiment designed to study the overall facial colour appearance of human faces. The same 80 real facial images as Experiments 1 and 2, including 40 Caucasian images and 40 Chinese images, were used in this experiment. Before the experiment, eighty corresponding colour patches were generated using MATLAB (an example shows in Figure 3.8). Each uniform colour patch has the same shape as the corresponding facial image, and the initial colour was the average facial colour calculated from all the pixels in the facial skin area of that image. The process of calculation can be found in Section 4.2.1 and an example of the facial skin area is shown in Figure 4.1(the non-black area).

The experiment was carried out in a dark room. The observers performed all the operations in a self-compiled MATLAB program on display. After 3-min dark adaption, the observers were given the experimental instructions and started a training session before the formal experiment to learn how to adjust the colour of the patch colour through keyboard control before the experiment. Figure 3.7 shows the instruction page the observers saw on display before the formal experiment.
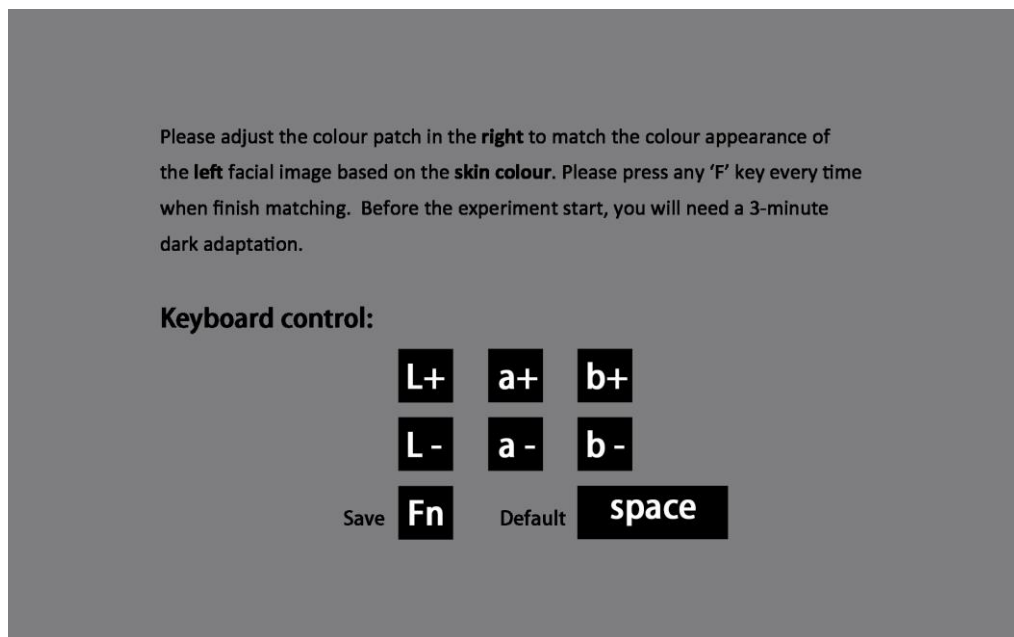
Figure 3.7 The experimental instructions page of Experimental 3.

Figure 3.8 shows an example of the experimental interface of the formal experiment. During the experiment, there was a facial image on the left of

the screen and a same-shape colour patch on the right. The observers were asked to adjust the patch colour without any time limit until they produced a match between the overall colour appearance of the facial image and the colour of the colour patch based on the skin colour.
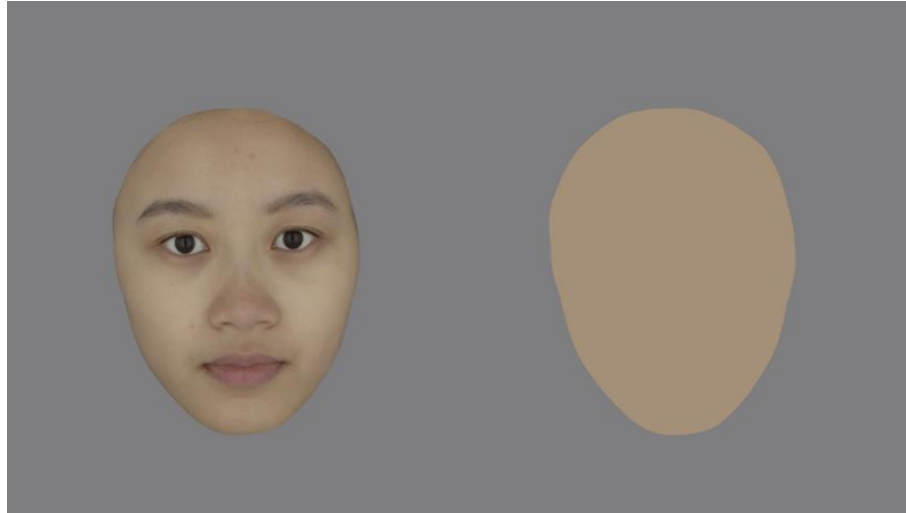
Figure 3.8 An example of the Experimental 3 interface.
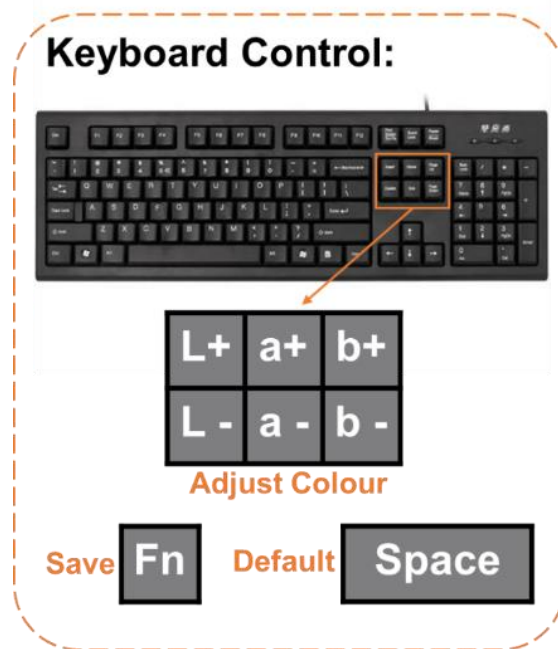
Figure 3.9 Keyboard control for patch colour adjustment.

All the colour control of the patch was achieved through the keyboard. As Figure 3.9 shows, six adjacent keys outlined on the keyboard were used to adjust the patch colour along the L*, a* and b* dimensions in CIELAB colour space, roughly altering facial lightness, redness, and yellowness, respectively. The step of adjustment was 0.5 units for each time pressing the keys of any colour dimensions. The function key was used to save the match

results of the present face and show the next face. The space key was used to restore the default patch colour.

After completing one match and pressing the Fn key, the data would be saved and the next set of facial images and colour patches would appear on display automatically. The 80 facial images were shown in random order and 10 randomly selected facial images were repeated to test the intra-observer agreement (repeatability). Besides, a reference facial image was also matched, which was not used in this experiment but was used in Experiment 4. In total, each observer matched the overall colour appearance of 90 facial images and an additional reference image.

3.4.1.2 Observers and observer variations

Forty-four observers with normal colour vision took part in the experiment, including 21 Caucasians (4 males; mean age $\pm$ SD = 25.67 $\pm$ 4.39) and 23 Chinese (11 males; mean age $\pm$ SD = 26.64 $\pm$ 3.87). They were either students or members of staff from the University of Leeds.

The observer variability including the inter- and intra- observer variability was evaluated in terms of the mean colour difference from the mean (MCDM). The inter-observer variability was evaluated per facial image by calculating the colour difference ΔE^*_{ab} between the matched colour of an individual observer and the mean-matched colour over all observers and then taking the mean. The mean inter-observer MCDM values calculated for all 80 faces was 2.61 ΔE^*_{ab} unit (Caucasian faces: 2.58 ΔE^*_{ab} unit, Chinese faces: 2.24 ΔE^*_{ab} unit). The Intra-observer variability was assessed by calculating the MCDM, regarding each observer's MCDM for the 10 repeated facial images and then averaging over all observers. All the observers showed high repeatability with an MCDM of less than 2 ΔE^*_{ab} units except for Observer 21 and 25 (the intra-observer variability was 2.21 ΔE^*_{ab}, and 2.23 ΔE^*_{ab} respectively). Thus, the observation data from these two observers were removed for further data analysis. The mean intra-observer MCDM values of the rest 42 observers were 1.13 ΔE^*_{ab} unit. These results of inter- and intra-observer variability indicated both the high consistency and repeatability of observers.

3.4.2 Experiment 4

3.4.2.1 Study design

Experiment 4 was a colour appearance scaling experiment designed to study perceptual facial whiteness, redness, and yellowness. The perceptual difference between facial colour and uniform patch colour was also

considered. The technique of magnitude estimation was used in the experiment. The observers were asked to scale the perceptual whiteness, redness, and yellowness of both the faces and the uniform colour patches. The same 80 real facial images used in Experiments 1, 2, and 3 were also used in this experiment together with 80 corresponding uniform colour patches. Each uniform colour patch has the same shape as the corresponding facial image, and the patch colour was the overall colour appearance of the corresponding facial image based on the colour matching results of Experiment 3.

The experiment was carried out in a dark room and observers had a 3-min dark adaption before the experiment. The observers performed all the operations in a self-compiled MATLAB program on display. The experiment was separated into two sessions, one to scale the facial images and the other to scale the uniform colour patches. The order of the two sessions was fully randomized for different observers. To ensure the observers had a clear concept of the experiment, they were asked to view all the images quickly, then were given the experimental instructions and started a training session before both sessions during the formal experiment. Figure 3.10 shows the instruction page for the facial image session and Figure 3.11 shows the instruction page for the uniform patch session. Note that the instruction texts on the top would disappear once the formal experiment started. The verbal instructions were also given to the observers: 'You will be shown a series of testing facial images/face-shaped uniform colour patches in random order. Your task will be to scale what whiteness, redness, and yellowness you see based on your perception of the skin/patch colour. Please use the reference as a standard, which had a whiteness/redness/yellowness of 50, to scale the perceptual whiteness/redness/yellowness for each testing face/ colour patch'.

During the facial image session, two facial images, including a reference face and a testing face, are shown on the display side by side. The reference face was an average face derived from 20 Caucasian faces and 20 Chinese faces and its whiteness, redness, and yellowness were all set to have a value of 50. Observers were asked to estimate the facial whiteness, redness, and yellowness of the testing face in comparison to the reference face respectively. An estimation scale of 0-100 was used where 0 represented the least white, red, or yellow face, and 100 represented the most white, red, or yellow face.
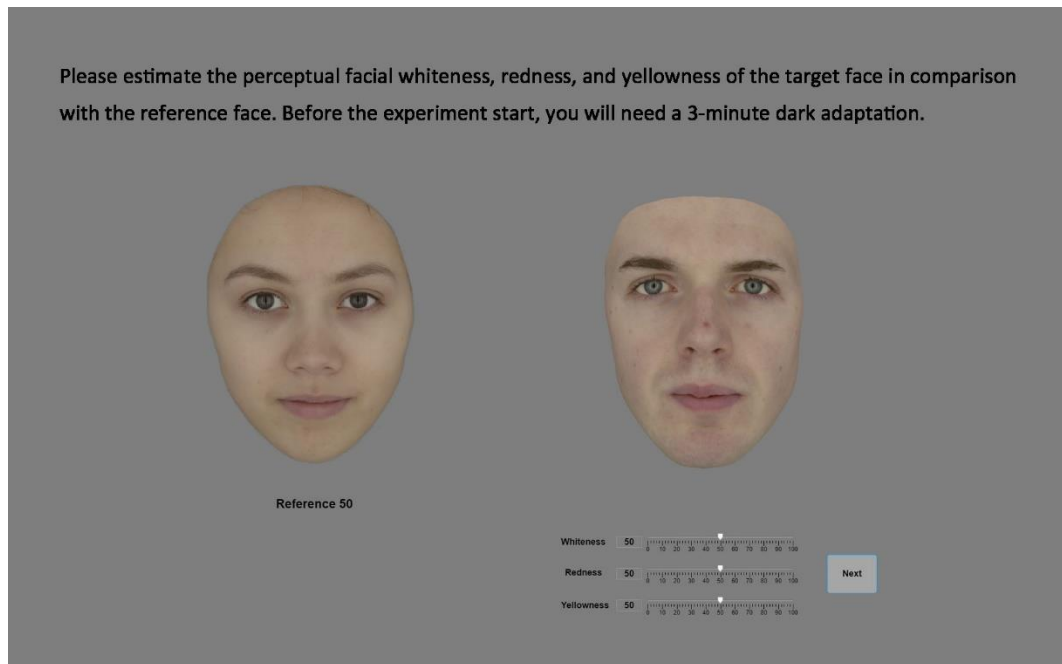
Figure 3.10 The experimental instructions page of the facial image session.

During the uniform patch session, two colour patches, including a reference patch and a testing patch, are shown on the display side by side. The colour of the reference patch was the overall colour appearance of the reference face, and the colour of the testing patch was the overall colour appearance of the corresponding facial image. The overall colour appearance of both the reference face and all the testing faces was obtained by colour matching in Experiment 3. The reference patch was set to have a value of 50 for whiteness, redness, and yellowness. Observers were asked to estimate the whiteness, redness, and yellowness of the testing colour patch in comparison to the reference colour patch. An estimation scale of 0-100 was used where 0 represented the least white, red, or yellow patch, and 100 represented the most white, red, or yellow patch.

Please estimate the perceptual whiteness, redness, and yellowness of the target patch in comparison with the reference patch. Before the experiment start, you will need a 3-minute dark adaptation.

Reference 50

Whiteness 50

Redness 50

Next

Yellowness 50
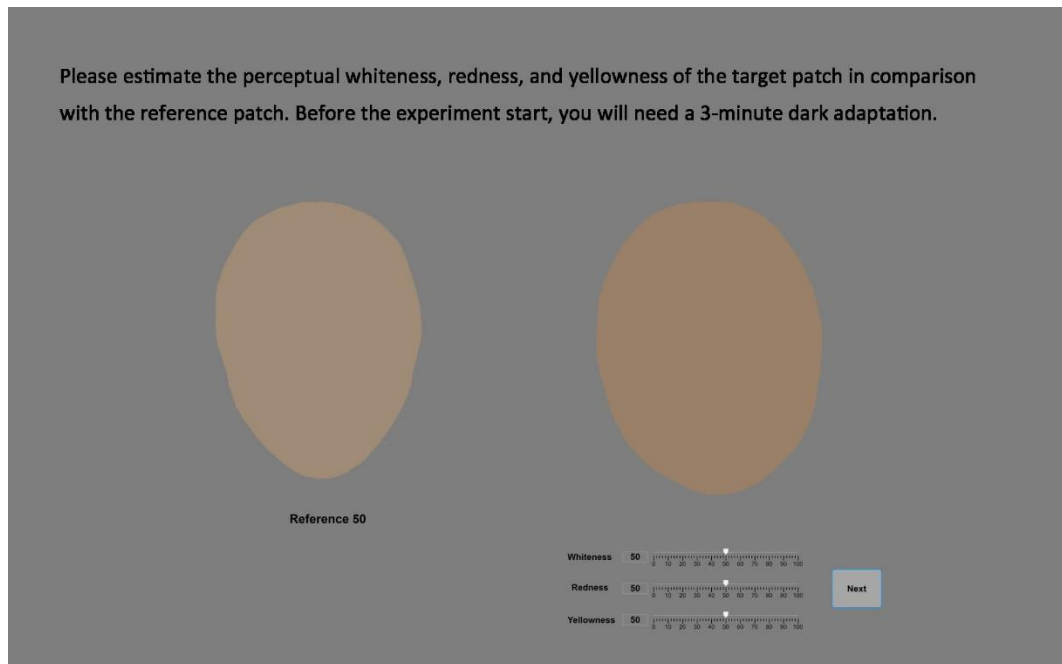
Figure 3.11 The experimental instructions page of the uniform patch session.

After completing the estimation of one face or one patch and clicking the Next button, the data would be saved for further data analysis and the next testing face or patch would appear on display automatically. During the facial image session, the 80 facial images were shown in random order and 10 randomly selected facial images were repeated to test the consistency of each observer. The uniform patch session included 80 corresponding patches as well we 10 randomly repeated patches. In total, each observer scaled the colour appearance of 90 facial images and 90 corresponding uniform colour patches.

3.4.2.2 Observers and observer variations

A panel of 43 observers took part in the experiment, including 23 Caucasians (7 males; mean age $\pm$ SD = 24.65 $\pm$ 4.61) and 20 Chinese (9 males; mean age $\pm$ SD = 25.70 $\pm$ 5.27). They were either students or members of staff from the University of Leeds. All the observers had normal colour vision according to the Ishihara test (Clark, 1924).

The coefficient of variation (CV) was used to verify the observer variation and the scaling factor k was set to one as the same scale was used in this experiment. The intra-observer variability was assessed by calculating the CV values between the two repeated judgements. The inter-observer variability was evaluated by computing the CV values between the individual observer scaling data and the mean scaling data of all observers. Table 3.3

summarized the resulting mean CV values covering all observers for the intra- and inter-observer variability.

Table 3.3 The intra- and inter-observer variability in terms of the CV values of face scaling and patch scaling for each of the three perceptual colour attributes.

Perceptual attributes	Inter-observer variability		Intra-observer variability	
	Face	Patch	Face	Patch
Whiteness	14.4	15.6	14.5	15.8
Redness	10.8	14.4	11.5	13.6
Yellowness	16.8	19.6	18.6	21.0

The intra- and inter-observer variability were close for both face and patch perception and for scaling all three perceptual attributes, which indicated that observers performed similarly in terms of within an individual observer and between observers. The largest mean CV values were seen for yellowness estimation and the redness perception had the smallest mean CV values, which was the same case for both face and patch. Thus, it could be assumed that observers might use different criteria in the evaluation of yellowness, whereas similar or common criteria in the perception of redness. For all three appearance attributes, smaller CV values were obtained for facial image scaling compared to uniform patch scaling, which suggested observers might find the skin colour was easier to scale, whereas the uniform patch colour was more difficult to scale.

3.5 Summary

Two groups of psychophysical experiments for accumulating the preference and appearance evaluation data were conducted. The details of experimental preparation and experimental procedures were described in this chapter and summarised below.

- The characteristics of the experimental displays were carefully investigated and the GOG model was used to perform the display colour characterisation for both the BenQ display and iPad display.
- The experimental materials, including 40 Caucasian facial images and 40 Chinese facial images, were selected from LLSD and then processed for Experiments 1, 3, and 4 based on the BenQ display model. The 40 Chinese facial images and another 60 Chinese facial

images were processed for Experiment 2 based on the iPad display model.

- The study design, experimental procedures, and psychophysical techniques were described in detail respectively for each of the four experiments.
- Observer variations for each experiment were also assessed using different measures according to the techniques used in the experiments.

Chapter 4 Average skin colour and facial preference

4.1 Overview

In the previous chapter, Experiment 1, a cross-cultural experiment on facial colour preference was introduced (Section 3.3.1). In this chapter, the raw data collected in Experiment 1 was used to test the role of average facial colour in preference judgements. The average skin colour, in terms of the L*, a*, b* values, of each facial image used in Experiment 1 was analysed (Section 4.2), statistical analysis (Section 4.3) was performed to elucidate the associations between average facial skin colour and perceptual ratings. Whether these associations are modulated by observer or image ethnicity was also investigated in this chapter.

4.2 Image analysis of the average facial skin colour

4.2.1 Image analysis

The mean colour specification, in terms of CIELAB coordinates (L*, a*, b*), of the 80 test facial images (40 Chinese and 40 Caucasian) were calculated as the overall mean of each pixel in the facial area, excluding the mouth, nose, eyes, and eyebrows, as shown in Figure 4.1. The areas other than the facial skin were masked manually for each of the images and the calculations were performed in MATLAB.

Figure 4.1 An example of the facial area (the non-black area) used to calculate the average facial colour.

4.2.2 The distribution of the average facial colours

Figure 4.2 shows the distribution of the mean facial colours of the test facial images in a*b* plane and L*C* plane in CIELAB colour space. There are systematic mean differences in lightness and chromaticity between the two

ethnic groups. The Caucasian images (average L*=59.0, a*=8.3, b*=14.1) have, on average, higher lightness (L*) and lower yellowness (b*) compared to the Chinese images (average L*=55.0, a*=8.9, b*=16.9).
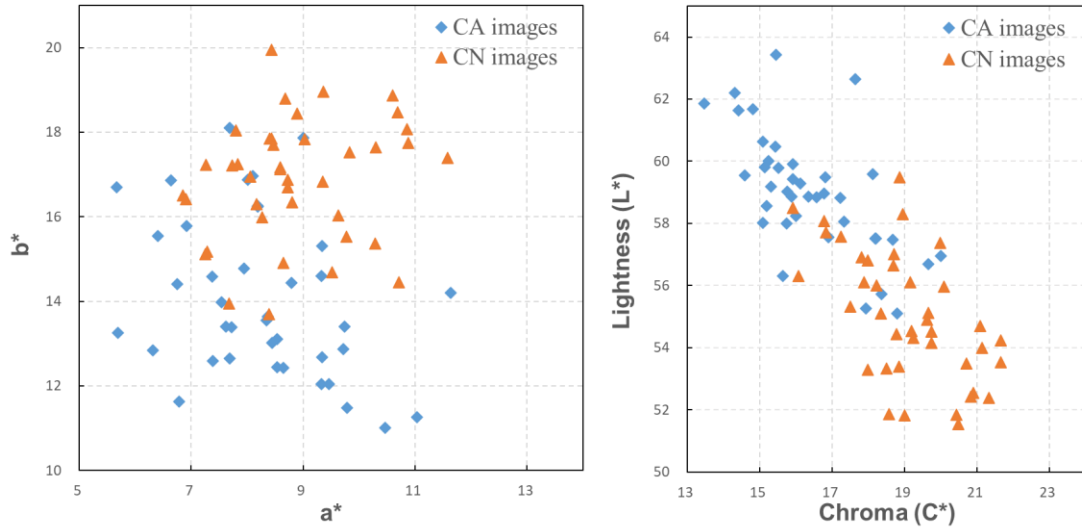
Figure 4.2 The distribution of the mean facial colours of the test facial images in CIELAB a*b* space (left) and L*C* space (right): ◆ Caucasian (CA), ▲ Chinese (CN).

4.3 Statistical analysis

The raw data collected from Experiment 1 was used to conduct the following data analysis. For each of the three dependent variables (DV - attractiveness, healthiness, perceived age) a linear mixed-effect model was set up with the following fixed effects: lightness (L*), redness (a*) and yellowness (b*) as continuous predictors and image ethnicity and observer ethnicity as categorical predictors, including random intercepts for both images and observers.

All linear mixed-effect models were implemented in lme4 R package (Bates et al., 2015). Deviation coding was used to convert both image ethnicity and observer ethnicity into deviation-coded factors (code '-0.5' for Caucasian images/observers and code '0.5' for Chinese images/observers) for testing the main effects of each model. For each DV, a simple and a full model was considered; the full model allowing for all interactions between colour (L*, a*, b*) and ethnicity (observer ethnicity, image ethnicity). The simple and full models were compared by the AIC and BIC values (Akaike, 1974; Schwarz, 1978; Burnham and Anderson, 2004). P values for the fixed effects in each linear mixed-effect model were calculated using F tests with type III sums of

squares and Satterthwaite's degrees-of-freedom approximation in the *lmerTest* R package (Kuznetsova et al., 2017). Significant interactions revealed in the tests were followed up with a further analysis of the simple effects for each subgroup. In addition, Pearson's correlation coefficients (two-tailed) were also used to test the associations of the perceptions of all three attributes for both sets of observers.

4.4 The effect of average skin colour on facial preference judgements

For all perceptual attributes, a linear mixed-effect model with and without interactions was first evaluated. In all cases, a model allowing for interactions outperformed the model without interactions, as shown in Table 4.1. While the BIC for age weakly favoured the model with no interaction, all the AIC strongly preferred the model with interactions, as did the significant likelihood ration test. Therefore the analysis of the model with interactions was reported in below.

Table 4.1 Model comparisons: mixed models with and without interactions for all three attributes.

Model	npar	AIC	BIC	logLik	deviance	χ^2	χ^2df	P
DV=attractiveness								
no interaction	9	9725	9781	-4854	9707			
+ interactions	19	9534	9651	-4748	9496	211.03	10	<0.001[***]
DV=healthiness								
no interaction	9	9471	9526	-4726	9453			
+ interactions	19	9344	9460	-4653	9306	147.1	10	<0.001[***]
DV=age								
no interaction	9	20013	20068	-9997	19995			
+ interactions	19	20002	20119	-9982	19964	31.09	10	<0.001[***]

[*]P≤0.05, [**] P≤0.01, [***]P≤0.001.

Table 4.2 shows all the main effects of the linear mixed-effect model for facial attractiveness. For attractiveness, neither the average skin colour (L*,a*,b*) nor observer ethnicity were significant, but image ethnicity, interactions between image ethnicity and lightness (p = 0.017) and interactions between observer ethnicity and L*/a*/b* are significant (p = 0.001/ p < 0.001/ p < 0.001). The interactions indicated that the effect of lightness on attractiveness was different when faces of different ethnicity

were viewed and the effects of L*/a*/b* on attractiveness were different for Caucasian observers and for Chinese observers.

Table 4.2 Linear mixed effects model estimates of fixed effects, their SE, t-value, lower (2.5%) and upper (97.5%) confidence intervals and P-values for attractiveness.

Fixed effects	Estimate	SE	*t*-value	2.5% CI	97.5% CI	P-value
(Intercept)	2.346	4.465	0.525	-6.513	11.206	0.601
L*	0.021	0.058	0.359	-0.094	0.136	0.720
a*	-0.113	0.085	-1.328	-0.282	0.056	0.188
b*	0.106	0.061	1.739	-0.015	0.228	0.086
Im	-20.703	8.929	-2.319	-38.419	-2.987	**0.023***
Ob	-1.531	1.780	-0.860	-5.020	1.958	0.390
Im:Ob	-6.266	3.546	-1.767	-13.218	0.686	0.077
L*:Im	0.284	0.116	2.448	0.054	0.515	**0.017***
L*:Ob	0.077	0.023	3.353	0.032	0.123	**0.001*****
a*:Im	0.187	0.170	1.102	-0.150	0.525	0.274
a*:Ob	-0.166	0.034	-4.901	-0.232	-0.099	**<0.001*****
b*:Im	0.182	0.122	1.487	-0.061	0.425	0.141
b*:Ob	-0.089	0.024	-3.646	-0.136	-0.041	**<0.001*****
L*:Im:Ob	0.079	0.046	1.707	-0.012	0.169	0.088
a*:Im:Ob	0.137	0.068	2.030	0.005	0.270	**0.042***
b*:Im:Ob	0.050	0.049	1.024	-0.046	0.145	0.306

*P≤0.05, ** P≤0.01, ***P≤0.001. Im=Image ethnicity, Ob=Observer ethnicity.

Table 4.3 shows the results of the main effects in the full perceived healthiness model. For perceived healthiness, there was no significant main effect of average skin colour, but a significant effect of image ethnicity. Significant interactions were found between image ethnicity and lightness (p=0.024) and between observer ethnicity and a* and b*, respectively (p<0.001; p<0.001). Similar to facial attractiveness, the effect of lightness on perceived healthiness was different to faces of different origins and the effect of redness/yellowness on healthiness is different for the two groups of observers.

Table 4.3 Linear mixed effects model estimates of fixed effects, their SE, t-value, lower (2.5%) and upper (97.5%) confidence intervals and P-values for healthiness.

Fixed effects	Estimate	SE	t-value	2.5% CI	97.5% CI	P-value
(Intercept)	4.829	4.998	0.966	-5.088	14.745	0.337
L*	-0.015	0.065	-0.231	-0.144	0.114	0.818
a*	-0.100	0.095	-1.055	-0.289	0.088	0.295
b*	0.093	0.068	1.357	-0.043	0.229	0.179
Im	-20.374	9.995	-2.039	-40.205	-0.544	**0.045***
Ob	2.141	1.725	1.241	-1.242	5.523	0.215
Im:Ob	-4.515	3.434	-1.315	-11.248	2.218	0.189
L*:Im	0.300	0.130	2.306	0.042	0.558	**0.024***
L*:Ob	0.027	0.022	1.208	-0.017	0.071	0.227
a*:Im	0.133	0.190	0.697	-0.245	0.510	0.488
a*:Ob	-0.141	0.033	-4.309	-0.205	-0.077	**<0.001***
b*:Im	0.137	0.137	1.002	-0.134	0.409	0.320
b*:Ob	-0.147	0.024	-6.237	-0.193	-0.101	**<0.001***
L*:Im:Ob	0.060	0.045	1.339	-0.028	0.147	0.181
a*:Im:Ob	0.102	0.065	1.563	-0.026	0.231	0.118
b*:Im:Ob	0.034	0.047	0.726	-0.058	0.126	0.468

*P≤0.05, ** P≤0.01, ***P≤0.001. Im=Image ethnicity, Ob=Observer ethnicity.

For estimated age, as shown in Table 4.4, there were significant main effects of redness (a*), image ethnicity, and the interaction between the ethnicity of the image and observer. Significant main effects of interactions between colorations and ethnicity included the interaction between lightness and image ethnicity (p=0.015), the interaction between redness and image ethnicity (p=0.036) and the three-way interaction of L*:Im:Ob (p=0.035). Facial lightness/redness had different effects on the perceived age of Caucasian faces and Chinese faces, and the influence of lightness on age perception also depended on the ethnicity of the observer.

Table 4.4 Linear mixed effects model estimates of fixed effects, their SE, t-value, lower (2.5%) and upper (97.5%) confidence intervals and P-values for estimated age.

Fixed effects	Estimate	SE	*t*-value	2.5% CI	97.5% CI	P-value
(Intercept)	23.676	14.317	1.654	-4.731	52.082	0.102
L*	-0.112	0.186	-0.603	-0.482	0.257	0.548
a*	0.550	0.273	2.017	0.009	1.091	**0.047**[*]
b*	0.237	0.196	1.210	-0.152	0.626	0.230
Im	70.588	28.617	2.467	13.805	127.37	**0.016**[*]
Ob	4.551	8.080	0.563	-11.290	20.392	0.573
Im:Ob	31.712	16.046	1.976	0.254	63.170	**0.048**[*]
L*:Im	-0.926	0.372	-2.487	-1.665	-0.187	**0.015**[*]
L*:Ob	-0.072	0.104	-0.693	-0.277	0.132	0.488
a*:Im	-1.165	0.545	-2.136	-2.247	-0.083	**0.036**[*]
a*:Ob	0.294	0.153	1.922	-0.006	0.593	0.055
b*:Im	-0.602	0.392	-1.536	-1.380	0.176	0.129
b*:Ob	-0.142	0.110	-1.294	-0.358	0.073	0.196
L*:Im:Ob	-0.440	0.209	-2.106	-0.849	-0.030	**0.035**[*]
a*:Im:Ob	-0.133	0.306	-0.434	-0.732	0.467	0.665
b*:Im:Ob	-0.313	0.220	-1.425	-0.744	0.118	0.154

[*]P≤0.05, [**] P≤0.01, [***]P≤0.001. Im=Image ethnicity, Ob=Observer ethnicity.

To further understand the interactions above and reveal the simple effects of L*, a* and b* within each set of image and observer, parameter estimates for the fixed effects within each subgroup were computed from the three linear mixed-effect models, as shown in Table 4.5.

Table 4.5 Parameter estimates of the simple effects in the linear mixed-effect models.

Fixed effects	CA observers	CN observers
DV = attractiveness		
CA images		
Model		
L*	β=-0.140，P=0.149	β=-0.102，P=0.291
a*	β=-0.090，P=0.507	**β=-0.324，P=0.018[*]**
b*	β=0.072，P=0.425	β=-0.041，P=0.647
CN images		
Model		
L*	β=0.105，P=0.133	**β=0.221，P=0.002^{**}**
a*	β=0.029，P=0.790	β=-0.068，P=0.538
b*	**β=0.229，P=0.010^{**}**	β=0.166，P=0.059
DV = healthiness		
CA images		
Model		
L*	β=-0.164，P=0.131	β=-0.166，P=0.124
a*	β=-0.071，P=0.638	β=-0.263，P=0.083
b*	β=0.106，P=0.292	β=-0.058，P=0.567
CN images		
Model		
L*	β=0.106，P=0.170	**β=0.163，P=0.037[*]**
a*	β=0.011，P=0.930	β=-0.079，P=0.519
b*	**β=0.226，P=0.021[*]**	β=0.097，P=0.318
DV = age		
CA images		
Model		
L*	β=0.277，P=0.381	β=0.424，P=0.180
a*	**β=0.952，P=0.033[*]**	**β=1.312，P=0.004^{**}**
b*	β=0.531，P=0.074	β=0.545，P=0.066
CN images		
Model		
L*	β=-0.429，P=0.060	**β=-0.721，P=0.002^{**}**
a*	β=-0.146，P=0.684	β=0.081，P=0.821
b*	β=0.085，P=0.763	β=-0.213，P=0.451

[*]P≤0.05, ^{**} P≤0.01, ^{***}P≤0.001. DV=dependent variable.

When Chinese observers rated Chinese faces, an increase in lightness was strongly associated with greater attractiveness (p = 0.002); for Caucasian faces, a decrease in redness predicted greater attractiveness (p = 0.018). Caucasian Observers associated an increase in yellowness with higher

attractiveness, but only when viewing Caucasian facial images (p = 0.010). Lighter skin was associated with greater healthiness but only when Chinese observers rated Chinese images (p = 0.037). Caucasian observers associated an increase in yellowness with healthiness (p = 0.021) when viewing facial images of Chinese. This association was driven by the Caucasian images: an increase in redness was associated with an older perceived age for Caucasian images when viewed by Caucasian (p = 0.033) or Chinese Observers (p = 0.004). Lightness was a strong predictor for perceived youthfulness when Chinese observers rated Chinese faces (p = 0.002).

4.5 Cultural differences on the associations between the three perceptual attributes

Ratings of attractiveness and healthiness are highly correlated across both image and observer ethnicities (Table 4.6, also see Figure 4.3) but are negatively correlated with estimated age. The latter negative correlations are highly significant for Chinese observers. The strongest negative correlations are observed when Chinese observers rate Chinese image, consistent with interactions between ethnicity and skin coloration cues.

Table 4.6 The Pearson Correlation Coefficients of age, healthiness, and attractiveness scores for the Caucasian (CA) and Chinese (CN) observers.

	CA images	CN images	Overall images
CA observers			
Attractiveness-Healthiness	0.912***	0.946***	0.929***
Attractiveness-Age	-0.343	-0.354	-0.351*
Healthiness-Age	-0.293	-0.295	-0.298
CN observers			
Attractiveness-Healthiness	0.881***	0.927***	0.893***
Attractiveness-Age	-0.632***	-0.828***	-0.730***
Healthiness-Age	-0.651***	-0.818***	-0.726***

*P≤0.05/18, ** P≤0.01/18, ***P≤0.001/18. N = 40, 40, 80 for CA, CN and overall images, respectively. All p-values were Bonferroni-corrected.
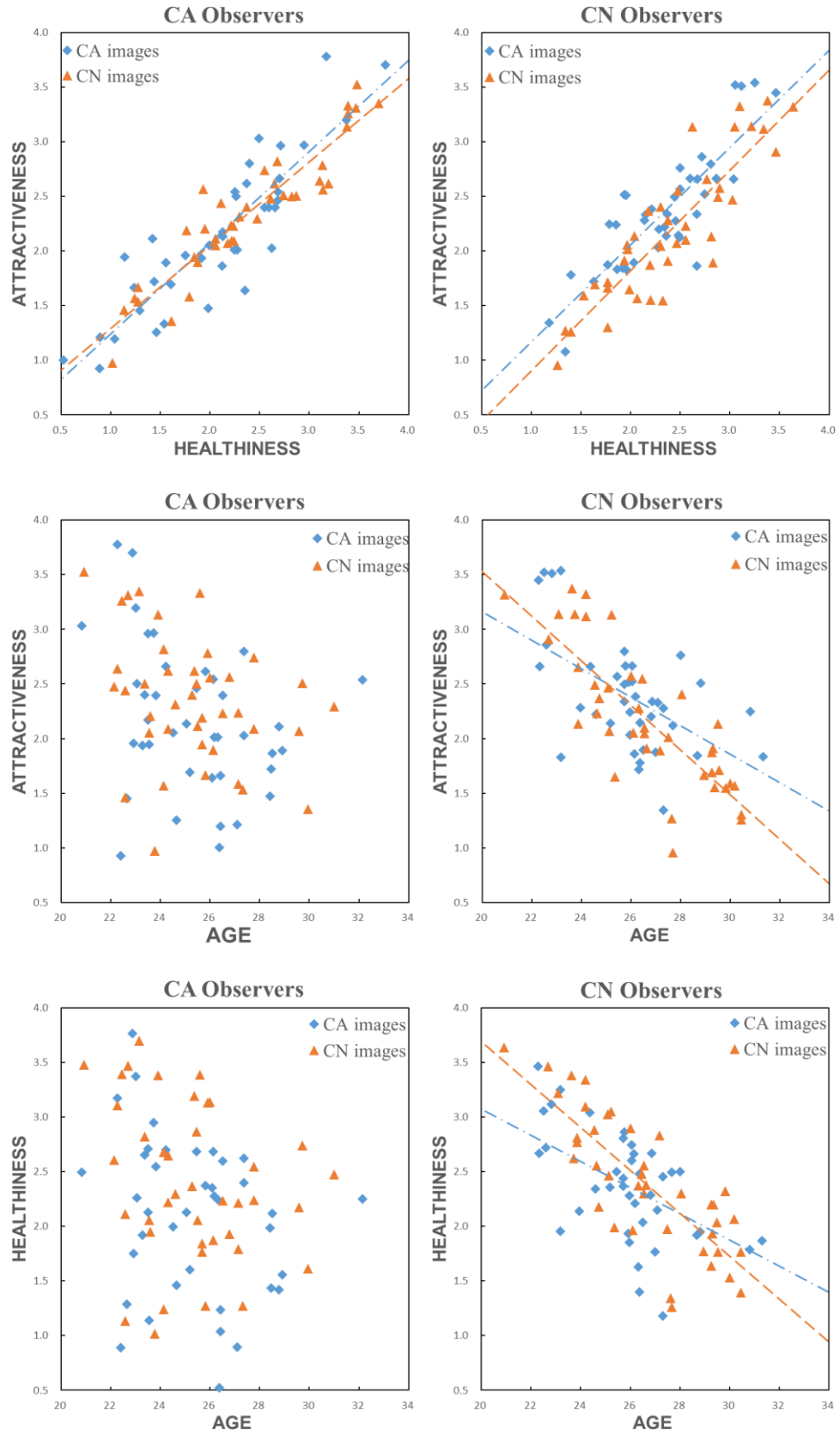
Figure 4.3 Associations between attractiveness, healthiness and age ratings for Caucasian observers (left column) and Chinese observers (right column). The regression lines were drawn for the significant correlations at Bonferroni adjusted α.

Comparison of the results of the Caucasian observers and the Chinese observers showed that there were differences between the two sets of responses for attractiveness, healthiness and age. For Caucasian observers, there was high correlation between attractiveness and healthiness but no correlations between attractiveness and age, and between healthiness and age. This applied to both Caucasian and Chinese facial images. Caucasian observers had highly correlated perceptions for facial attractiveness and healthiness, but they wouldn't link these two attributes with the estimated age. For Chinese observers, however, attractiveness was correlated significantly and positively with healthiness, age correlated significantly but negatively with both attractiveness and healthiness. When Chinese observers judged their own ethnicity images, the three attributes were very highly correlated with each other (positively or negatively), which indicated that, for Chinese observers, the three perceptions, attractiveness, healthiness, and age, could be considered consistent. Perceptual difference or aesthetic difference was found in this part.

4.6 Discussions

4.6.1 The use of the realistic skin models

In total, 80 calibrated, high-resolution, non-manipulated images of real human faces were used in Experiment 1 as representatives of the complexion and skin colour variation of real, Caucasian and Chinese, human faces. The ethnic difference of skin colour was larger along the lightness and yellowness dimension and smaller in redness, which was consistent with the skin colour variation of the two ethnic groups (Xiao et al., 2017). Although the observers were asked to make judgements based only on skin colour, it was difficult to ignore the role of certain facial features, for example, the eyes, the nose, the lips and the mouth. In our study, this real variation was covered, at least partially, by using a set of 80 images of real faces, 40 Caucasian and 40 Chinese.

4.6.2 The role of average skin colour (L*, a*, b*) in preference judgement is more limited than previously thought

The most robust positive associations were found between facial skin lightness (L*) and attractiveness, healthiness, and youthfulness, but only when Chinese observers judged facial images of their own ethnicity. These associations between ratings and skin lightness are grounded in known

physical changes: a decrease of one L* unit is equivalent to a 10-year increase in age in female Chinese skin (Huixia et al., 2012). In contrast to our study, Han et al. (Han et al., 2018) found an effect of lightness on attractiveness/perceived health for both, Chinese and Caucasian faces, whereas in this study significant associations was only found when Chinese faces were rated. A possible explanation for this discrepancy is that Chinese skin is characterised by a lower average L* value, and crucially, by a smaller variation in lightness compared to the skin of Caucasians (Xiao et al., 2017). It is speculated the smaller variability in L* for Chinese faces leads to a more informative and more reliable lightness cue.

The association between an increase in skin yellowness (b*) and perceived attractiveness and healthiness is likely to reflect the Caucasian preference for 'tanned' skin. Skin yellowness as a significant predictor for perceived health is consistent with previous studies (Whitehead et al., 2012). Whether skin yellowness is associated with physical health is controversial and not clear so far.

No evidence was found that facial skin redness is positively associated with perceived attractiveness, healthiness or youthfulness (Tables 4.5), in contrast to previous reports with Caucasian (Stephen, Coetzee, et al., 2009; Stephen, Law Smith, et al., 2009; Stephen et al., 2011; Stephen et al., 2012) and Chinese observers (Han et al., 2018; Tan and Stephen, 2019). The results of this study were, however, consistent with recent studies using a large image data base of real female faces that did not find any association between facial redness and objective health measures, neither any positive association between redness and attractiveness (Cai et al., 2019). Other studies employing non-manipulated real facial images, found a weak positive association between skin yellowness and facial attractiveness, but skin redness as a mediator showed a small but negative association with facial attractiveness (Appleton et al., 2018).

Broadly speaking, studies that have reported a strong positive association between average skin colour and attractiveness or healthiness, have involved colour-manipulated facial images (Stephen, Coetzee, et al., 2009; Stephen, Law Smith, et al., 2009; Lefevre and Perrett, 2015; Pazda et al., 2016). More recent studies, including this study, using non-manipulated facial images, have failed to show these strong associations (Foo, Simmons, et al., 2017; Jones, 2018; Appleton et al., 2018; Cai et al., 2019). This discrepancy could be due to the methodological differences, including the magnitude of colour changes (e.g. redness in excess of 10 a* units) and

colour shifts being applied uniformly across the face. While in the real human faces, skin colour doesn't change uniformly over the entire face and the magnitude is much smaller (the range of redness values covered by different faces is only about 6 a* units, Figure 4.2). Crucially, skin colour manipulations were often restricted to the CIELAB dimensions, a*, b*, L*, whereas in the natural skin colour universe, skin colour dimensions are highly correlated (Xiao et al., 2017). Thus, it could be speculated that previous experiments may have overestimated these associations by using skin colour manipulations well beyond the gamut found in non-manipulated images as well as changing the coloration of the entire face instead of specific areas.

4.6.3 The role of other facial colour cues

Since non-manipulated images of real faces were used, average facial colour co-varied necessarily with other facial colour cues when preference judgement was made based on the skin colour. The high observer consistency both within and across ethnicities (Table 3.2) suggest that observers may also rely on additional facial colour cues in their judgments. The current part of work was designed to estimate the strength of association between mean skin coloration (L*, a*, b*) and the perceptual ratings of perceived health, attractiveness, and perceived age; it is not allowed to estimate the contribution of skin coloration relative to the contribution of other facial colour cues, such as skin colour variation, localised skin colour or other feature colour contrasts. A further analysis in the next chapter would be useful to include all the other facial colour cues and evaluate the relative contribution of both the average skin colour and other facial colour characteristics.

4.6.4 The perceptual difference between the three facial attributes among Caucasian and Chinese observers.

Perceived attractiveness and healthiness showed significantly high correlations among both Caucasian and Chinese observers. Cultural differences emerged however, when age judgements were taken into consideration. Caucasian observers thought that both a younger face and older face could have an attractive and healthy appearance. For Chinese observers, younger faces meant a healthier look and were more attractive. These results were consistent for observers judging both own-ethnicity faces and other-ethnicity faces. Neotenous faces were perceived as more attractive and healthier by Chinese observers. Such ethnic differences in objective aesthetic criteria should be considered in many applications of

preferred skin colour reproduction such as the various surgeries performed for aesthetic reasons (Gao et al., 2018).

4.7 Summary

This chapter focused on the role of overall facial colour appearance and how it relates to attractiveness, healthiness, and perceived age. A summary of the analysis and major findings are given below:

- The average skin colour (L*, a*, b*) of all the facial images used in Experiment 1 was analysed. Statistical analysis was conducted to assess the role of the average L*, a*, b* in facial preference judgement and reveal the cultural difference in the perception of attractiveness, healthiness, and age based on facial colour appearance.
- Evidence was found that observers of both ethnicities make use of the average facial skin colour (L*, a*, b*) to rate attractiveness, healthiness, and perceived age, but the utilisation of these cues is more subtle than previously thought.
- Crucially, those skin coloration cues are not universal and are utilised differently within the Chinese and Caucasian ethnic groups, reflecting different aesthetic preferences in eastern and western cultures.
- These results contributed to the growing body of work demonstrating the importance of skin colour manipulations within an evolutionary meaningful parameter space, ideally using realistic skin models based on physical parameters.

Chapter 5 Various colour characteristics and facial preference

5.1 Overview

In the last chapter, the limited role of average skin colour in preference evaluation in real faces was discussed. In this chapter, various facial colour characteristics were quantified and their ranges and variations in Chinese and Caucasian populations were analysed (Section 5.2). Using the data from Experiment 1, separate analyses were carried out for each ethnic group to examine the role of colour in predicting the preference rating of their own faces. Techniques from machine learning were implemented to perform a comprehensive assessment of the relative importance of various facial colour characteristics that contribute to facial attractiveness, perceived healthiness, and perceived age. Three different classes of colour cues were compared (Section 5.5) and the most important colour predictors were identified (Section 5.6). Cultural differences between Chinese and Caucasian are further explored taking into account a range of facial colour characteristics.

5.2 Image analysis of facial colour characteristics

5.2.1 Image analysis

Based on the literature review, nineteen facial colour characteristics from the three classes in below were considered as potential important colour cues for facial preference judgements and were analysed for each of the 80 facial images in Experiment 1. All the areas of interest shown in Figure 5.1 were selected manually for each image and all the calculations were performed in MATLAB.
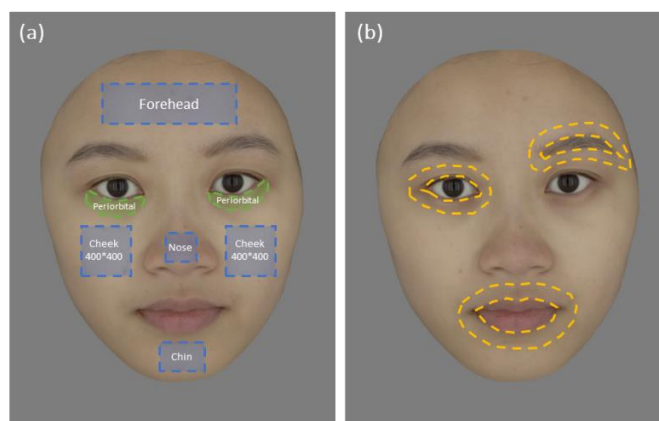
Figure 5.1 An example of the facial image showing the areas that selected for calculating facial colour characteristics. (a) Areas of interest used to calculate local skin colour and skin colour variation; (b) Areas of the features and the surrounding skin used to calculate facial colour contrasts.

Average facial colour and local skin colour. Apart from the average facial skin colour (L*, a*, b*), the local skin colour of cheek redness, a*, and periorbital lightness, L*, were also taken into account considering the study of Jones et al. (Jones et al., 2016). They were calculated as the overall mean of each pixel within the selected areas as shown in Figure 5.1(a).

Skin colour variation. To access the facial skin colour variation, the mean colour difference from the mean (MCDM) was adopted, a measure commonly used to describe colour variation for a set of data points in CIELAB space using the following equation (Nadal et al., 2011; Berns, 2019),

$$MCDM = \frac{\sum_{i=1}^{N}[(L_i^* - \overline{L^*})^2 + (a_i^* - \overline{a^*})^2 + (b_i^* - \overline{b^*})^2]^{1/2}}{N}$$ **Equation 5.1**

Here, MCDM was used to evaluate skin colour variation of any target facial areas, where L_i^*, a_i^*, and b_i^* are the CIELAB coordinates for the ith pixel of the area, $\overline{L^*}$, $\overline{a^*}$, and $\overline{b^*}$ are the average CIELAB coordinates of the facial area and N is the number of pixels within the area. As outlined in Figure 5.1(a), the MCDM of the forehead, cheek, nose, and chin areas was calculated and the grand mean of the MCDM values of the four parts was then obtained to represent the skin colour variation over the whole facial area. Both the skin colour variation of the whole facial area (MCDM) and the cheek (MCDM-Cheek) were analysed in this study. The smaller the value of the MCDM, the smaller the colour difference and the more even/homogeneous the skin colour distribution is.

Facial colour contrast. The adapted version of the Michelson contrast and the CIELAB colour differences (ΔE) were used to describe facial colour contrast between three facial features (eyes, eyebrows, and mouth) and their surrounding skin Figure 5.1(b). The adapted Michelson contrast of the three dimensions (L*, a*, b*) was considered, as defined by the following equation,

$$C_{\text{Feature}} = \left| \frac{A_{\text{Skin}} - A_{\text{Feature}}}{A_{\text{Skin}} + A_{\text{Feature}}} \right|$$ **Equation 5.2**

where A_{Skin} is the respective CIELAB coordinates (L*, a*, b*) of the surrounding facial skin and A_{Feature} is the respective CIELAB coordinates (L*, a*, b*) of the facial features (eyes, eyebrows, and mouth). Meanwhile, the CIELAB colour differences (ΔE) between the three facial features and their surrounding skin were also calculated and the facial colour contrast was defined by the following equation,

$$\Delta E = [(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2]^{1/2}$$

Equation 5.3

where L_1^*, a_1^*, and b_1^* are the CIELAB coordinates of the facial features, and L_2^*, a_2^*, and b_2^* are the CIELAB coordinates of their surrounding skin area. For both C_{Feature} and ΔE, the bigger the value, the larger the facial colour contrast is.

5.2.2 Variation in facial colour characteristics across Caucasian and Chinese images

All the facial colour characteristics were quantified in CIELAB colour space. Table 5.1 and Figure 5.2 shows all the parameters measured for the forty Caucasian (CA) faces and the forty Chinese (CN) faces. CA and CN faces differ in various facial colour characteristics. The mean values and standard deviations for each group can be found in Table 5.1, together with the results of a two-sample t-test (P values) for the difference between the two ethnic datasets. All colorimetric characteristics between the two ethnic sample differ statistically from each other (P<0.05), except for the cheek redness (cheek-a*) and the skin colour variations (MCDM-cheek and MCDM).

Table 5.1 Descriptive statistics for the facial colour characteristics in CA and CN facial images (*t*-test values for which P>0.05 are shown in bold).

	Mean (SD)		*t*-test
	CA	CN	
L*	60.63 (2.06)	55.92 (2.10)	<0.001
a*	8.10 (1.26)	8.72 (1.16)	0.025
b*	15.25 (2.06)	18.55 (1.50)	<0.001
Cheek-a*	9.91 (1.70)	10.04 (1.91)	**0.741**
Periorbital-L*	54.77 (2.87)	53.07 (2.11)	0.003
MCDM-Cheek	2.48 (0.32)	2.40 (0.31)	**0.237**
MCDM	2.58 (0.21)	2.51 (0.22)	**0.169**
Eyes-C-L*	0.15 (0.03)	0.18 (0.03)	<0.001
Eyes-C-a*	0.21 (0.09)	0.13 (0.07)	<0.001
Eyes-C-b*	0.14 (0.07)	0.21 (0.04)	<0.001
Eyes-△E	15.02 (2.78)	18.25 (2.47)	<0.001
Brows-C-L*	0.16 (0.06)	0.13 (0.05)	0.010
Brows-C-a*	0.06 (0.05)	0.13 (0.08)	<0.001
Brows-C-b*	0.06 (0.04)	0.18 (0.04)	<0.001
Brows-△E	16.32 (5.18)	13.67 (4.17)	0.014
Mouth-C-L*	0.12 (0.02)	0.10 (0.02)	0.001
Mouth-C-a*	0.41 (0.06)	0.33 (0.06)	<0.001
Mouth-C-b*	0.14 (0.05)	0.17 (0.04)	0.002
Mouth-△E	17.84 (2.64)	15.26 (2.13)	<0.001
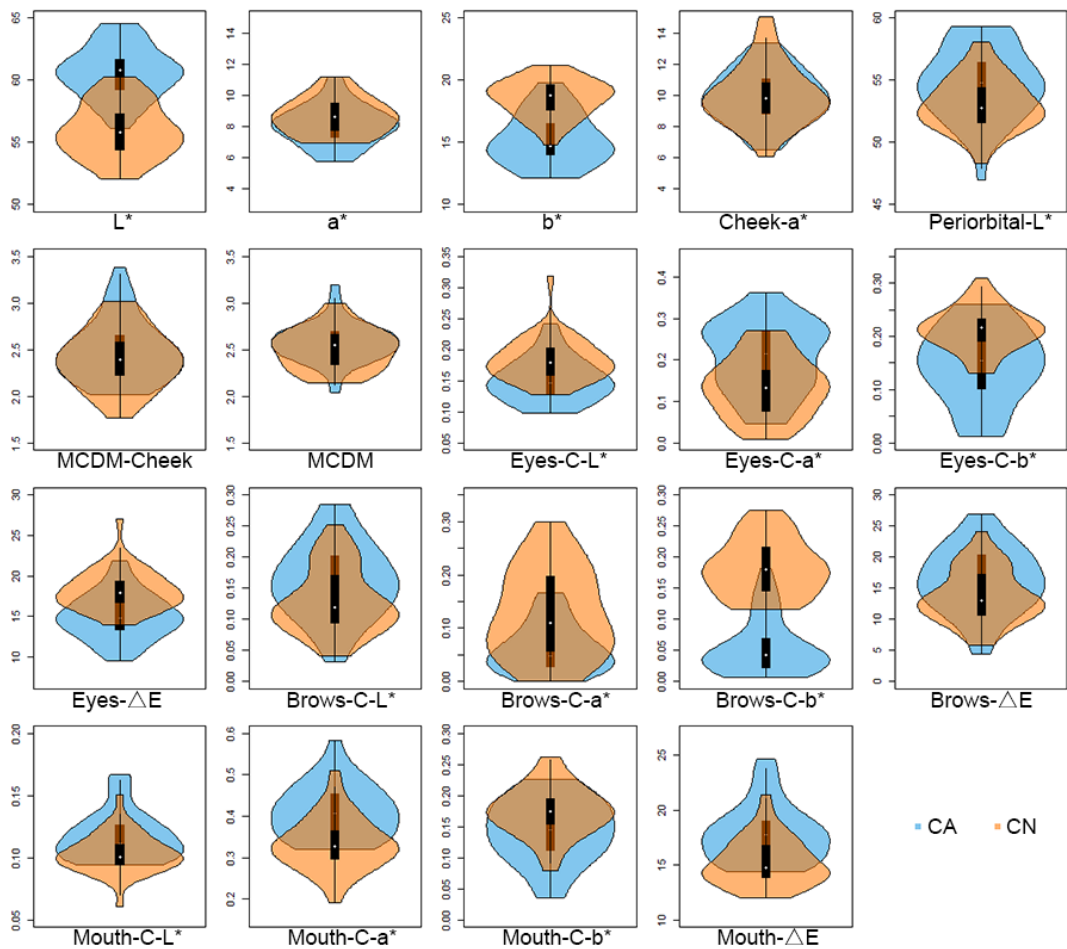
Figure 5.2 Violin plots showing range and variation of facial colour characteristics in CA and CN facial images. White points indicate medians, black rectangles represent interquartile ranges.

5.3 Statistical analysis and modelling techniques

Considering in most cases in daily life observers are making preference judgements of faces of their own ethnicity, this part of analysis focused on the 'within-race' effect. Separate analyses were carried out for each ethnic group to examine the colour variables that might predict each preference rating. Thus, the observation data collected in Experiment 1 was separated into two dataset, the Caucasian dataset, which is the preference ratings of the Caucasian images judged by the Caucasian observers, and the Chinese dataset, which is the preference ratings of the Chinese images judged by the Chinese observers.

As the internal consistency in the ratings of attractiveness, healthiness, and age for both the Caucasian dataset and Chinese dataset is very high, ranging from Cronbach's α = .90 to Cronbach's α = .96 for Caucasian

dataset and from Cronbach's α = .92 to Cronbach's α = .96 for Chinese dataset across the attributes (Table 3.2). Ratings were averaged across all observers to create a score for each face on each preference attributes before correlation analysis and modelling from the face level colour traits. All the colour predictors were z-standardised prior to the following analysis as prepared for the cross-validated regressions.

The Pearson Correlation Coefficient (two-tailed) was first used to identify correlations between each colour characteristic and facial attractiveness, healthiness, and visual age rated by the observers, for the Caucasian and Chinese datasets, respectively. The results will be described in Section 5.4.

To compare the role of the three classes of colour characteristics (average facial colour and local skin colour, skin colour variation, facial colour contrasts) in predicting the preference of real human faces, and identify their relative importance, techniques of cross validation from machine learning were implemented. The analysis was done by the caret package in R (Kuhn, 2008). Cross-validated linear regression models (5-fold cross validation with 50 repeats) were used to compare the predictive power of the three different classes. The data was split into a 4-fold training dataset for model estimation and a 1-fold testing dataset for predictive accuracy test, so that the problem of overfitting could be avoided by testing the model with the new testing data rather than the old training data. Moreover, the process was repeated 50 times with different random splits of the data. The model's overall predictive fit was assessed by the mean RMSE (root mean square error) over all splits as RMSE represents the difference between predicted values from the model and observed values from the experiments. RMSE has the same unit as the scale used in experiment and is not inflated by the number of predictors, compared to other statistics such as R^2. The results will be reported in Section 5.5.

Finally, the Elastic Net Regression (Hastie et al., 2001; Zou and Hastie, 2005) was used to evaluate the role of all facial colour characteristics and their relative importance in determining facial preference by simultaneously entering them into one regression model. The Elastic Net Regression is a linear regression which shrink predictors to reduce overfitting through Regularisation and meanwhile perform variable selection by setting the coefficients of uninformative parameters to zero. The models have two hyperparameters which could be tuned to optimize the model fit, α, which controls the degree to which the model shrinks coefficients, and λ, which determines how aggressively coefficients are set to zero. Cross validation

was also implemented here to first generate the optimal combination of the two model hyperparameters, α and λ, with maximized fit (minimized RMSE) and then test the model fit with the optimal α and λ by the mean RMSE over all splits. The performance of each combined model represented by the mean RMSE and the relative importance of different facial colour characteristics in the model represented by the absolute β values will be reported in Section 5.6.

5.4 Zero-order correlations between facial colour characteristics and each facial preference

The results of the correlations between facial colour characteristics and each of the three preference ratings are shown in Figure 5.3 and the complete correlation matrix of preference ratings and facial colour characteristics can be found in Table 5.2 at the end of this section.

Facial attractiveness. Facial colour characteristics were linked differently with facial attractiveness by the Caucasian (solid bars) and the Chinese observers (dashed bars). In the Caucasian dataset, facial attractiveness was positively correlated with facial yellowness (b*, $p<0.05$) and b* contrast around the mouth (mouth-C-b* $p<0.05$), but negatively with L* contrast around the mouth (mouth-C-L*). In the Chinese dataset, facial attractiveness was positively correlated with facial lightness (L*, $p<0.01$), a* contrast around the mouth (mouth-C-a*, $p<0.001$), and colour difference around the mouth (mouth-$\triangle$E, $p<0.01$), which may also result from the a* contrast considering the high correlation between a* contrast and $\triangle$E around the mouth (r=0.859, $p<0.001$). Chinese facial attractiveness was negatively correlated with facial redness (a*), both skin colour variation (MCDM-cheek and MCDM), and a* contrast around the brows (brows-C-a*).

Perceived healthiness. The attractiveness ratings and healthiness ratings are highly correlated for both groups (r>0.9, $p<0.001$ in Table 4.6), thus colour cues utilized for healthiness perception were somewhat similar to those for attractiveness judgements. For the Caucasian dataset, perceived healthiness was positively correlated to facial yellowness (b*, $p<0.05$) and b* contrast around the mouth (mouth-C-b*, $p<0.05$), but negatively correlated to overall lightness (L*, $p<0.01$), periorbital lightness (periorbital-L*, $p<0.05$), and overall skin colour variation (MCDM, $p<0.05$). For the Chinese dataset, perceived healthiness was positively correlated to facial skin lightness (L*, $p<0.01$), colour contrast around the eye and the mouth (eyes-C-b*, $\triangle$E, $p<0.05$; mouth-C-a*, $\triangle$E, $p<0.01$). Perceived healthiness for the Chinese

dataset was negatively correlated with facial redness (a*, p<0.05) and a* contrast around the brows (brows-C-a*, p<0.01).
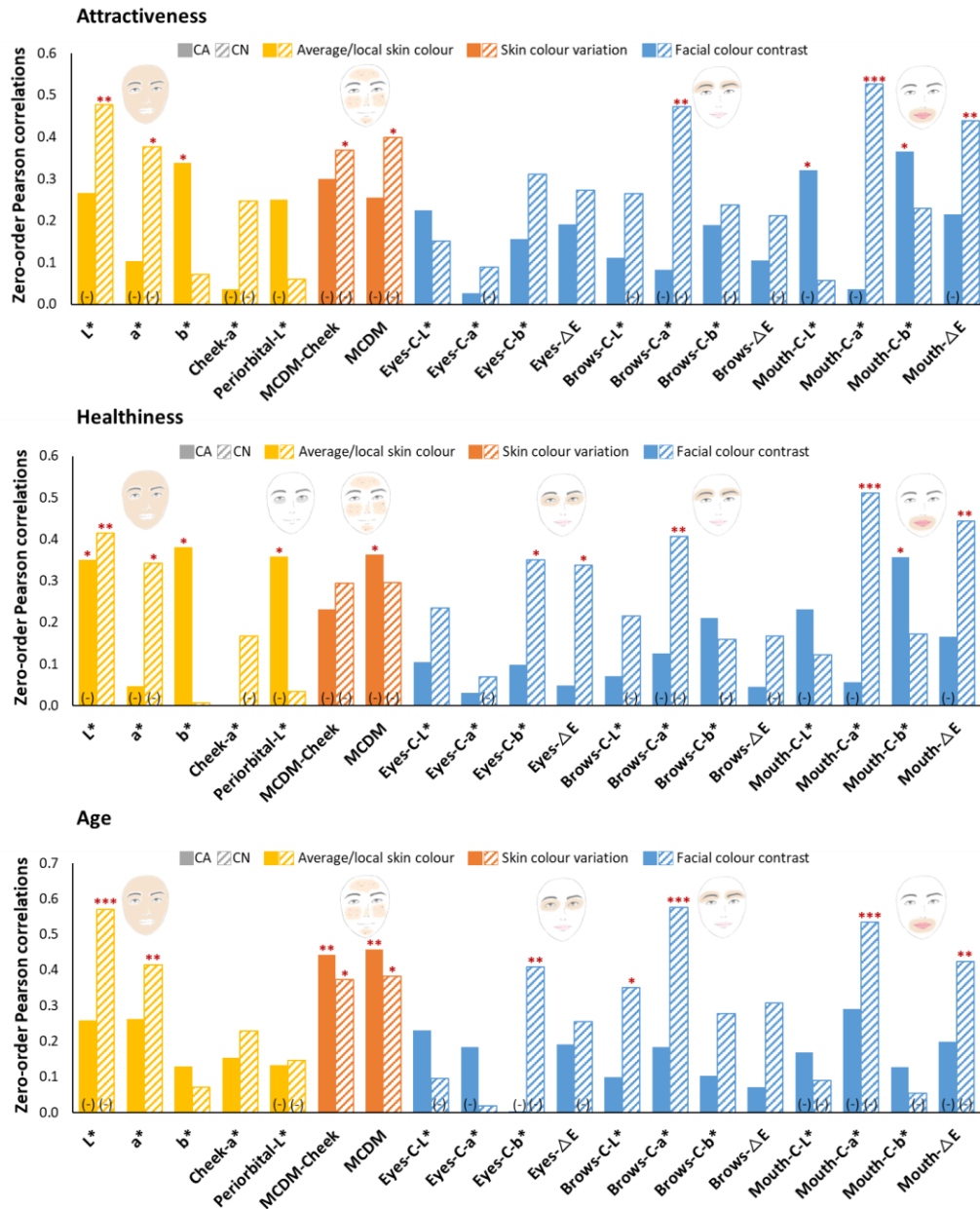
Figure 5.3 The Pearson Correlations between each facial colour characteristic and each facial preference attributes: attractiveness (top), healthiness (middle), and age (bottom). Each bar chart represents the correlation coefficient (left darker bar chart: CA; right lighter bar chart: CN); all the negative coefficients are marked with (-) at the bottom of the bar charts; Asterisks above the bar charts indicate the statistical significance of each relationship: *p≤0.05, ** p≤0.01, ***p≤0.001.

Perceived age. For the Caucasian dataset, perceived age was only significantly and positively associated with skin colour variation (MCDM-cheek and MCDM, both p<0.01), which means larger variation in Caucasian skin colour was linked to older visual age. For the Chinese dataset, in

addition to skin colour variation (MCDM-cheek and MCDM, both p<0.05), perceived age was also positively correlated with facial redness (a*, p<0.01), colour contrast around the brows (brows-C-L*, a*, p<0.05). In addition, it was negatively correlated with facial lightness (L*, p<0.001), colour contrast around the eye and mouth (eyes-C-b*, p<0.01; mouth-C-a*, p<0.001; mouth-$\triangle$E, p<0.01).

The colour difference, $\triangle$E, around the three facial features (eyes, brows, mouth-$\triangle$E) was excluded in the analysis of the next two section since it has originated from one of the separate colour contrast channels (L*, a*, or b*) for both groups according to the correlations in Table 5.2 (r>0.86, p<0.001).

Table 5.2 Zero-order Pearson correlations between facial colour characteristics and preference ratings. CN results are below the diagonal and CA results are above the diagonal. The different coloured boxes show correlation at p< 0.001, p<0.01 and p<0.05.

	Attractiveness	Healthiness	Age	L*	a*	b*	Cheek-a*	Periorbital-L*	MCDM-Cheek	MCDM	Eyes-C-L*	Eyes-C-a*	Eyes-C-b*	Eyes-ΔE	Brows-C-L*	Brows-C-a*	Brows-C-b*	Brows-ΔE	Mouth-C-L*	Mouth-C-a*	Mouth-C-b*	Mouth-ΔE
Attractiveness		0.912***	-0.343*	-0.267	-0.103	0.338*	-0.036	-0.250	-0.300	-0.254	0.224	0.027	0.156	0.190	0.111	-0.083	0.190	0.104	-0.321*	-0.037	0.365*	-0.216
Healthiness	0.927***		-0.293	-0.351*	-0.046	0.381*	0.001	-0.359*	-0.232	-0.363*	0.104	0.030	0.098	0.049	0.070	-0.125	0.211	0.045	-0.232	-0.056	0.356*	-0.166
Age	-0.828***	-0.818***		-0.259	0.262	0.130	0.153	-0.133	0.442**	0.457**	0.231	-0.184	-0.003	0.192	0.098	0.184	0.102	0.070	-0.168	-0.290	0.128	-0.198
L*	0.477**	0.415**	-0.571***		-0.519***	-0.425**	-0.429**	0.715***	-0.318*	-0.296	-0.203	-0.096	-0.371*	-0.088	-0.180	0.209	-0.253	-0.073	0.261	0.651***	-0.382*	0.414**
a*	-0.377*	-0.342*	0.415**	-0.554***		-0.301	0.898***	-0.108	0.329*	0.193	0.061	0.124	0.024	0.024	-0.346*	-0.276	-0.130	-0.411**	-0.022	-0.616***	-0.269	-0.182
b*	0.071	0.006	0.072	-0.387*	0.153		-0.271	-0.570***	0.076	0.104	0.239	0.084	0.543***	0.228	0.428**	0.239	0.555***	0.411**	-0.276	-0.048	0.735***	-0.159
Cheek-a*	-0.247	-0.167	0.229	-0.406*	0.881***	0.109		-0.095	0.336*	0.181	-0.018	0.115	-0.020	-0.039	-0.416**	-0.286	-0.139	-0.471**	0.070	-0.436**	-0.240	-0.040
Periorbital-L*	0.060	0.034	-0.146	0.538***	-0.151	-0.458**	-0.156		-0.187	-0.189	-0.121	-0.127	-0.401*	0.015	-0.184	0.159	-0.257	-0.119	-0.018	0.222	-0.547***	0.066
MCDM-Cheek	-0.369*	-0.294	0.373*	-0.191	0.134	0.167	0.221	0.052		0.448**	0.117	-0.125	0.188	0.115	-0.031	0.256	0.196	-0.079	0.035	-0.305	0.017	-0.068
MCDM	-0.399*	-0.296	0.383*	-0.383*	0.124	0.194	0.066	-0.195	0.510**		0.227	-0.186	0.164	0.221	0.293	0.165	0.252	0.255	-0.109	-0.240	0.137	-0.171
Eyes-C-L*	0.151	0.235	-0.095	0.143	-0.279	-0.240	-0.020	-0.088	0.111	-0.036		-0.242	0.330*	0.971***	0.366*	0.309	0.493**	0.356*	-0.318*	-0.147	0.058	-0.250
Eyes-C-a*	-0.089	-0.069	0.018	-0.266	0.324*	0.270	0.303	-0.076	0.246	0.196	-0.196		0.465**	-0.159	-0.117	-0.198	-0.225	-0.117	0.241	-0.033	-0.141	0.190
Eyes-C-b*	0.310	0.350*	-0.409*	0.119	-0.164	0.174	0.017	-0.267	0.055	-0.150	0.239	0.470**		0.393*	0.262	0.089	0.315*	0.235	-0.177	-0.195	0.237	-0.153
Eyes-ΔE	0.273	0.338*	-0.254	0.337*	-0.383*	-0.269	-0.093	0.059	0.079	-0.120	0.963***	-0.151	0.356*		0.360*	0.376*	0.481**	0.365*	-0.332*	-0.107	0.004	-0.236
Brows-C-L*	-0.265	-0.215	0.351*	-0.181	-0.106	-0.195	-0.097	0.176	0.234	0.102	0.314*	-0.265	-0.334*	0.206		0.253	0.452**	0.990***	-0.300	0.079	0.367*	-0.215
Brows-C-a*	-0.472**	-0.406**	0.575***	-0.380*	-0.021	-0.292	-0.050	-0.042	0.169	0.180	0.248	-0.202	-0.428**	0.083	0.761***		0.206	0.278	-0.125	0.163	0.075	0.024
Brows-C-b*	-0.238	-0.159	0.277	0.012	-0.262	-0.421**	-0.196	0.361*	0.184	0.110	0.295	-0.201	-0.328*	0.234	0.681***	0.798***		0.420**	-0.265	-0.063	0.230	-0.173
Brows-ΔE	-0.211	-0.168	0.307	-0.085	-0.181	-0.220	-0.159	0.262	0.239	0.078	0.310	-0.297	-0.354*	0.223	0.989***	0.750***	0.736***		-0.267	0.159	0.356*	-0.158
Mouth-C-L*	0.057	0.122	-0.090	0.239	-0.250	-0.242	-0.162	0.106	0.043	0.036	0.369*	-0.278	0.002	0.360*	0.091	0.058	0.209	0.118		0.495**	-0.435**	0.915***
Mouth-C-a*	0.526***	0.511***	-0.535***	0.517**	-0.635***	-0.198	-0.522**	0.074	-0.514**	-0.360*	0.215	-0.500**	0.086	0.303	0.014	-0.079	0.125	0.065	0.316*		-0.170	0.750***
Mouth-C-b*	0.229	0.171	-0.054	-0.125	-0.174	0.496**	-0.189	-0.416**	-0.149	0.137	-0.018	-0.180	0.077	-0.029	-0.028	-0.066	-0.197	-0.032	-0.020	0.372*		-0.358*
Mouth-ΔE	0.438**	0.444**	-0.424**	0.468**	-0.452**	-0.082	-0.343*	0.050	-0.377*	-0.172	0.234	-0.492**	0.023	0.296	-0.026	-0.134	0.073	0.027	0.624***	0.859***	0.470**	

Column groups: average/local skin colour (L*, a*, b*, Cheek-a*, Periorbital-L*); skin colour variation (MCDM-Cheek, MCDM); facial colour contrast (Eyes-C-L* through Mouth-ΔE).

Legend: p<0.001, p<0.01, p<0.05

5.5 The separate model: comparisons of three classes of facial colour characteristics in determining facial preference.

Figure 5.4 shows each model's overall predictive fit was assessed by the mean RMSE (root mean square error) over all splits.
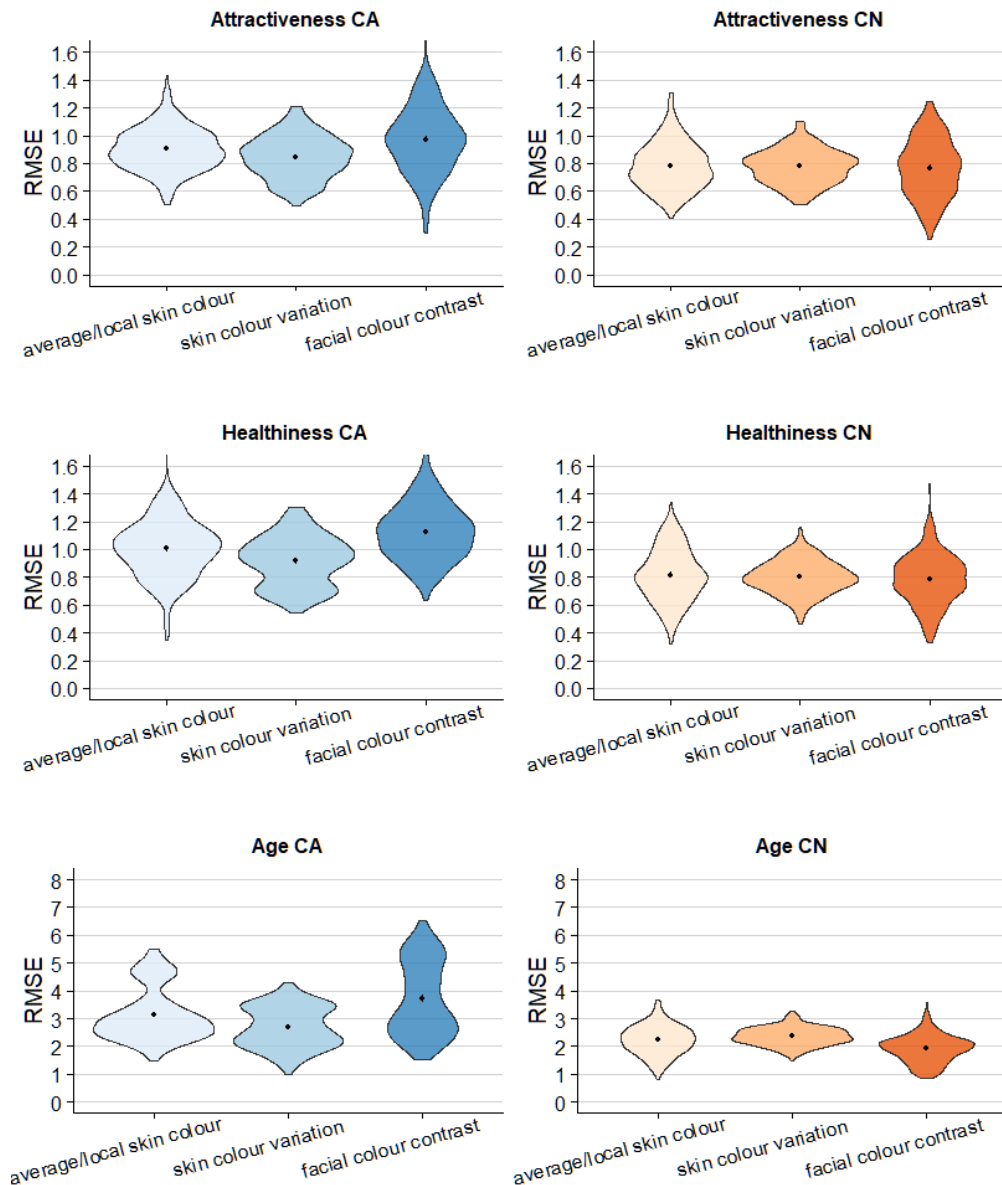
Figure 5.4 The model performance of the three classes of facial colour characteristics in predicting each facial preference attributes: attractiveness (top), healthiness (middle), and age (bottom). CA results are in the left column and CN results are in the right column. Black dots indicate the mean RMSE from 5-fold cross-validation with 50 repeats.

Facial attractiveness. For the Caucasian dataset, the model of skin colour variation showed the best predictive accuracy ($M_{RMSE} = 0.85$, SD_{RMSE}

= 0.15), followed by the model of average/local skin colour (M_{RMSE} = 0.90, SD_{RMSE} = 0.15), and facial colour contrast (M_{RMSE} =0.97, SD_{RMSE}=0.24). For the Chinese dataset, the three classes showed similar predictive accuracy (average/local skin colour: M_{RMSE} = 0.78, SD_{RMSE} = 0.17; skin colour variation: M_{RMSE} = 0.78, SD_{RMSE} = 0.12; facial colour contrast: M_{RMSE} = 0.77, SD_{RMSE} = 0.21).

Perceived healthiness.   For the Caucasian dataset, the model of skin colour variation showed the best predictive accuracy (M_{RMSE} = 0.91, SD_{RMSE} = 0.18), followed by the model of average/local skin colour (M_{RMSE} = 1.01, SD_{RMSE} = 0.18), and facial colour contrast (M_{RMSE} = 1.10, SD_{RMSE} = 0.20). For the Chinese dataset, the facial colour contrast showed the best predictive accuracy (M_{RMSE} = 0.78, SD_{RMSE} = 0.18), followed by the skin colour variation (M_{RMSE} = 0.80, SD_{RMSE} = 0.12), and the average/local skin colour (M_{RMSE} = 0.81, SD_{RMSE} = 0.19).

Perceived age.   For the Caucasian dataset, the model of skin colour variation showed the best predictive accuracy (M_{RMSE} = 2.69, SD_{RMSE} = 0.73), followed by the model of average/local skin colour (M_{RMSE} = 3.19, SD_{RMSE} = 0.93), and facial colour contrast (M_{RMSE} = 3.82, SD_{RMSE} = 1.31). For the Chinese dataset, the model of facial colour contrast showed the best predictive accuracy (M_{RMSE} = 1.92, SD_{RMSE} = 0.49), followed by the model of average/local skin colour (M_{RMSE} = 2.22, SD_{RMSE} = 0.53), and the skin colour variation (M_{RMSE} = 2.35, SD_{RMSE} = 0.32).

5.6 The combined model: role of facial colour characteristics in determining facial preference.

For each perceptual rating, the performance of the each combined (Elastic Net Regression) model was represented by the mean RMSE, and the relative importance of different facial colour characteristics in the model was represented by the absolute β values as shown in Figure 5.5 – Figure 5.7.
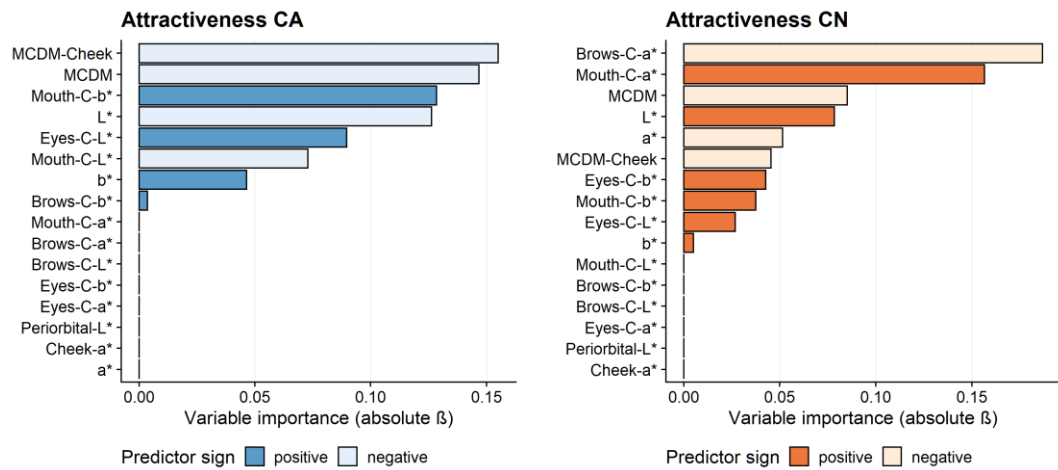
Figure 5.5 The relationship between different facial colour characteristics and facial attractiveness. CA results are in the left and CN results are in the right. Coefficients were derived from the elastic net model with 5-fold cross validation and 50 repeats.

Facial attractiveness. For the Caucasian dataset, the combined model predicted facial attractiveness within 0.84 point on a 7-point scale (M_{RMSE} = 0.84, SD_{RMSE} = 0.16). As shown in Figure 5.5, the skin colour variation (MCDM-cheek, $\bar{\beta}$ = -0.155 and MCDM, $\bar{\beta}$ = -0.147) were the strongest predictors, with less skin colour variation predicting higher facial attractiveness. The mouth colour contrast (mouth-C-b*, $\bar{\beta}$ = 0.128) and the facial lightness (L*, $\bar{\beta}$ = -0.126) were also relatively informative predictors, whereas the brows contrast ($\bar{\beta}$ = 0.003) was relatively uninformative. For the Chinese dataset, the combined model predicted facial attractiveness within 0.71 point on a 7-point scale (M_{RMSE}=0.71, SD_{RMSE}=0.14). The brows colour contrast (brows-C-a*, $\bar{\beta}$ = -0.187) and the mouth colour contrast (mouth-C-a*, $\bar{\beta}$ = 0.157) were the strongest predictors of facial attractiveness. The facial skin yellowness (b*, $\bar{\beta}$ = 0.005) was relatively less informative.
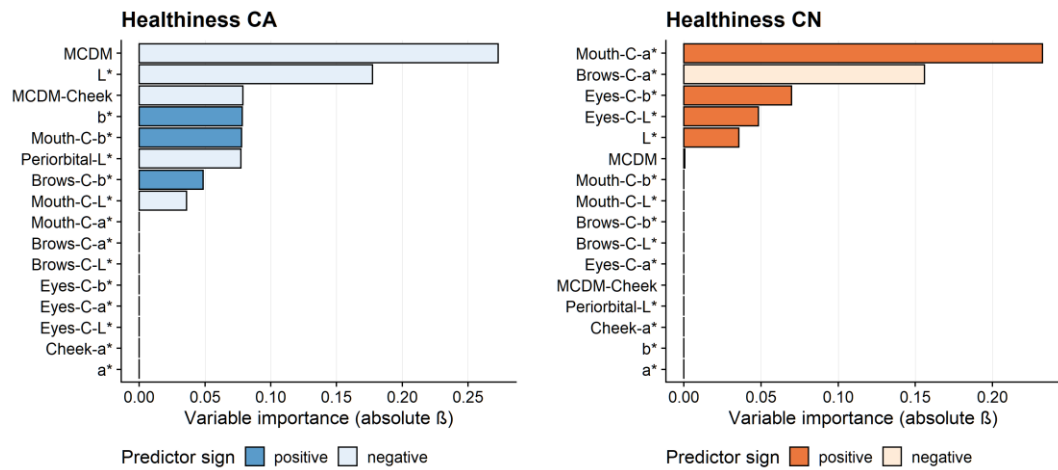
Figure 5.6 The relationship between different facial colour characteristics perceived healthiness. CA results are in the left and CN results are in the right. Coefficients were derived from the elastic net model with 5-fold cross validation and 50 repeats.

Perceived healthiness. For the Caucasian dataset, the combined model predicted perceived healthiness within 0.85 point on a 7-point scale (M_{RMSE} = 0.85, SD_{RMSE} = 0.16). As shown in Figure 5.6, the overall skin colour variation (MCDM, $\bar{\beta}$ = -0.273) and the facial lightness (L*, $\bar{\beta}$ = -0.177) were the strongest predictors, with less skin colour variation and lower skin lightness predicting higher perceived healthiness. The mouth luminance contrast (mouth-C-L*, $\bar{\beta}$ = -0.036) was relatively uninformative. For the Chinese dataset, the combined model predicted facial attractiveness within 0.73 point on a 7-point scale (M_{RMSE} = 0.73, SD_{RMSE} = 0.11). The mouth colour contrast (mouth-C-a*, $\bar{\beta}$ = 0.232) and the brows colour contrast (brows-C-a*, $\bar{\beta}$ = -0.156) were the strongest predictors of perceived healthiness. The overall skin colour variation (MCDM, $\bar{\beta}$ = -0.001) was relatively less informative.
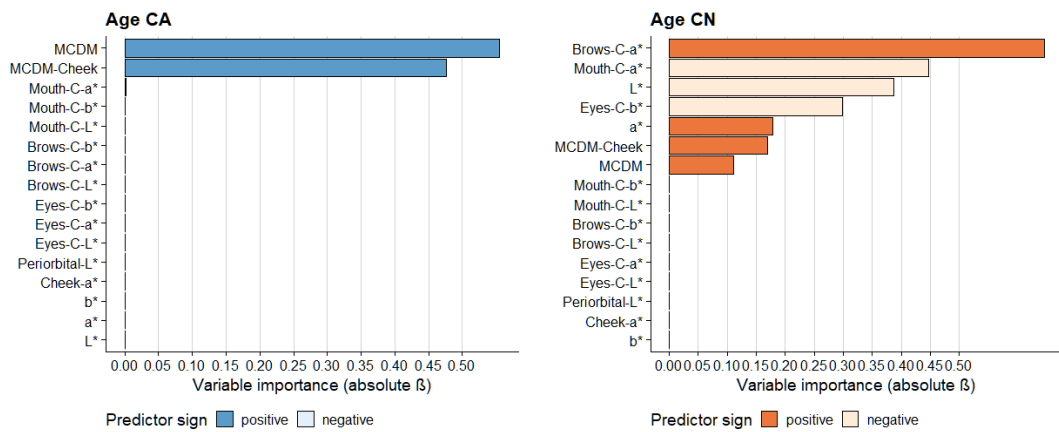
Figure 5.7 The relationship between different facial colour characteristics and perceived age. CA results are in the left and CN results are in the right. Coefficients were derived from the elastic net model with 5-fold cross validation and 50 repeats.

Perceived age.   For the Caucasian dataset, the combined model predicted perceived age within 2.88 years on a single-year step scale from 1 to 99 years (M_{RMSE} = 2.88, SD_{RMSE} = 0.99). As shown in Figure 5.7, the skin colour variation (MCDM, $\bar{\beta}$ = 0.555 and MCDM-cheek, $\bar{\beta}$ = 0.476) were the strongest predictors for perceived age, with a larger skin colour variation predicting a higher estimated age. The mouth colour contrast (mouth-C-a*, $\bar{\beta}$ = -0.002) was relatively uninformative. For the Chinese dataset, the combined model predicted perceived age within 1.83 years on a single-year step scale from 1 to 99 years (M_{RMSE} = 1.83, SD_{RMSE} = 0.34). Similar to the perceived healthiness, the brows colour contrast (brows-C-a*, $\bar{\beta}$ = 0.646) and the mouth colour contrast (mouth-C-a*, $\bar{\beta}$ = -0.446) were also the strongest predictors of perceived age. The facial lightness (L*, $\bar{\beta}$ = -0.387) and the eyes colour contrast (eyes-C-b*, $\bar{\beta}$ = -0.298) were also relatively informative predictors, whereas the overall skin colour variation (MCDM, $\bar{\beta}$ = 0.110) was relatively less informative.

5.7 Discussion

5.7.1 Colour predictors for facial attractiveness and perceived healthiness

For both Caucasian and Chinese observers, colour predictors of attractiveness and perceived healthiness are somewhat overlapping since these two perceptual ratings were highly correlated for both datasets (r=0.912 for CA dataset, r=0.927 for CN dataset, see Table 4.6).

For Caucasian observers, skin colour variation is the strongest predictor to rate attractiveness and perceived healthiness, and more evenly distributed skin colour with less variation is linked to enhanced facial attractiveness and perceived healthiness, which is consistent with previous studies (Fink and Matts, 2008; Fink et al., 2012). Compared to the skin colour variation, the averaged skin colour is less important and only the L*, b* were found to be predictors of attractiveness and healthiness. Corroborating the results of the last chapter, facial redness (a* or cheek-a*) is not an important predictor for Caucasian preference, and may be due to the small range of naturally occurring skin colour variation and thus the observers focus more on the other colour cues when rating the real facial images. Facial colour contrast did not emerge as an important predictor of preference in Caucasians, in contrast to previous studies (Stephen and McKeegan, 2010; Porcheron et al., 2013; Russell et al., 2016); only contrast around the mouth showed a limited role in attractiveness and perceived health. The reason for this is also likely to be the limited range of facial colour contrast in real faces without any applied cosmetics.

For Chinese observers, facial colour (a*) contrast (brows-C-a*, mouth-C-a*) is the most important predictor among different colour characteristics to judge both attractiveness and healthiness. Facial lightness (L*) is another consistent cue for Chinese people to judge facial preference, in line with the results of the last chapter (Table 4.5). The opposite preference for skin tanning between Chinese and Caucasians (Lefevre and Perrett, 2015) suggests the mainstream aesthetic difference between the two cultures. Skin colour variation also emerged as a predictor for Chinese observers but only when they judge facial attractiveness, with smaller variation in skin colour linked to enhanced facial attractiveness. Local skin colour does not emerge as a relevant predictor when all colour features are considered together.

5.7.2 Colour predictors for perceived age

Skin colour variation is found to be a predictor of perceived age in both the Caucasian and Chinese datasets (Figure 5.7).

Crucially, it is the only important colour cue for age perception of Caucasian observers judging own-ethnicity faces. Larger variation in facial/cheek skin colour is linked to older visual age. This is in agreement with the study of Nkengne et al. (Nkengne et al., 2008), which looked at the influence of various skin attributes (skin yellowness, skin texture, etc.) on the age perception of Caucasians and found that skin colour uniformity was the most important attribute. Chinese observers deploy the colour cues differently

from the Caucasian sample. Since all three perceptual ratings from Chinese observers are highly correlated (r>0.818, in Table 4.6), the significant colour predictors of perceived age are similar to the predictors of attractiveness and healthiness.

For Chinese observers, facial redness contrasts (brows and mouth) are the most important predictors for perceived age, but are deployed differently: brow and mouth contrasts are associated with a decrease and increase in youthfulness respectively. Skin lightness (L*) is the third informative cue for perceived age: a higher facial lightness is associated with youthfulness (younger visual age). Similarly to Caucasian observers, Chinese observers also rate more evenly distributed skin colour as younger. However, skin colour variation only plays a limited role compared to skin colour and contrast. These results reveal the importance of facial colour (a*) contrast for the Chinese observers. Using the same set of facial colour contrasts, Porcheron et al. investigated their relationship with the perceived age in Chinese subjects and also found the mouth a* contrast had significant and negative correlation with real age and the brows a* contrast had positive correlations with age (Porcheron et al., 2017).

5.7.3 Cultural difference between Caucasian and Chinese observers

As noted above, the use of different facial colour cues is ethnicity specific (Figures 5.5 - 5.7) and these results extended the findings in the last chapter on the ethnicity-specific use of the average facial skin colour. Moreover, the cultural differences include the opposite preference for facial lightness and the different importance of the three classes of colour traits in preference evaluation. The aesthetic difference between western and eastern culture might result from the development of multiple social and cultural factors over a long period of time. Meanwhile, the differential use of the facial colour cues could also stem from the different colorimetric parameters of the faces of the two ethnic groups (as shown in figure 5.2).

Generally, Chinese observers tend to utilise facial colour cues more effectively when evaluating facial preference (attractiveness, perceived healthiness, and visual age) compared to Caucasians, which is reflected in the higher number of significant correlations between perceptual ratings and colour characteristics in the Chinese dataset (Figures 5.3). The results of the separate models suggest that all the three classes of colour predictors show better predictive accuracy (smaller RMSE) in the Chinese models compared to the Caucasian models no matter which preference attribute is judged

(Figure 5.4). Most importantly, the Chinese combined models also give better predictive accuracy than the Caucasian combined models in all preference attributes, which predict attractiveness, healthiness, and visual age within 0.71 point, 0.73 point, and 1.83 years, respectively (the predictive accuracy is 0.84 point, 0.85 point, and 2.88 years for Caucasian model, respectively). These results suggest an important and novel aspect of the cultural difference between Caucasian and Chinese samples. Coetzee et al. investigated the role of facial shape cues and colour cues in attractiveness preference of White Scottish and Black South African people and found that Black South Africans rely heavily on colour cues while White Scottish use shape cues (Coetzee et al., 2014). Given that Asians were less influenced by some structural facial features than Caucasians (Cunningham et al., 1995), it is speculated that Caucasians may make facial preference judgements based on more structural facial features than colour cues while Chinese rely more heavily on facial colour cues.

5.7.4 The role of facial colour characteristics in preference evaluation of real faces

The results show the similar importance of all three classes of colour traits (average skin colour, skin colour variation, facial colour contrast) in determining facial preference judgements in real faces (Figure 5.4). Which colour characteristics are used depends on the preference attribute under consideration and also on the ethnic group (Figure 5.5 - 5.7).

As outlined in the Section 4.6.3, earlier studies using manipulated images have commonly found more significant relationships between the single manipulated colour cue and preference ratings (Fink et al., 2006; Stephen, Law Smith, et al., 2009; Scott et al., 2010; Jones et al., 2016; Porcheron et al., 2017). However, those colour manipulations in single dimensions have neglected the fact that facial colour preference is evaluated holistically in real situation based on various facial colour cues. When judging facial preference of real human faces, it may be beneficial to consider a wide range of facial colour cues simultaneously, hence allowing an estimate of the relative importance of the individual cues.

Though more recent studies have started to use non-manipulated images to study facial preference, and found much weaker associations between skin colour and facial preference (Nkengne et al., 2008; Foo, Simmons, et al., 2017; Jones, 2018; Tan et al., 2018; Appleton et al., 2018), none of them have considered all classes of different facial colour cues together. Foo et al. investigated skin colour (L*, a* and b*) and other structural facial features as

the preference predictors, and they concluded that skin colour did not predict attractiveness while facial yellowness played a limited role in predicting healthiness(Foo, Simmons, et al., 2017). Jones et al. also compared facial shape cues and colour cues in health perception using average facial L*, a*, and b*, and they found no role of skin colour as a short-term health cue (Jones, 2018). Tan et al. studied skin texture and colour in health perception and found homogenous skin texture and increased skin yellowness was positively associated with perceived health of Malaysian Chinese faces, however, facial colour contrast was not considered in their study which may also be an important predictor(Tan et al., 2018). Consistent with those studies that used non-manipulated images, the results showed that average skin colour (L*, a*, and b*) itself, as a single factor, is not a very strong predictor for facial preference evaluation but plays a limited role, especially for Caucasians' age perception. Given that different facial colour cues were utilised differently depending on the preference judgement at hand and the observers, a wide range of facial colour characteristics need to be studied at the same time to obtain a realistic estimate of the role of colour features for aesthetic preferences.

5.8 Summary

In this chapter, a comprehensive assessment of various facial colour characteristics that affect facial preference was provided. Colour predictors of facial attractiveness, perceived healthiness, and perceived age were studied in both Caucasian and Chinese samples. A summary of the analysis and major findings are given below:

- Various facial colour characteristics were calculated and these naturally occurring variations across Caucasian and Chinese faces were analysed.
- A moderate role for colour characteristics in determining facial preference was revealed.
- Although the average skin colour of facial areas plays a limited role, together with colour variation and contrast, there are stronger links between colour and facial preference than previously revealed.
- Different facial colour cues are found to be utilised by different observers according to the different preference attributes they are accessing. Interestingly, Chinese observers tend to rely more heavily on colour cues to judge facial preference than Caucasian observers.

- The results highlighted the importance of examining various facial colour cues simultaneously to characterise the role of colour predictors in facial preference evaluation and demonstrated the large cultural difference between Caucasian and Chinese populations.

Chapter 6 Analytical tool for facial attractiveness modelling

6.1 Overview

In the last chapter, attempts were made to examine the relationship between various colour cues and preference judgements in Caucasian and Chinese samples. The relative importance of different colour predictors was revealed. For numbers of industry applications, there is a strong need to predict facial preference. It is often desirable to achieve good predictive accuracy with fewer explanatory variables. Due to the high number of candidate variables and their correlations, predicting attractiveness from various colour cues could be difficult and may depend on the modelling methods.

Taking the Chinese dataset as an example, this chapter provided a complete analytic framework regarding attractiveness modelling and focused on the comparison of different multivariate regression techniques. The attractiveness ratings collected in Experiment 1 were used as the training dataset for model estimation. To avoid the issue of overfitting, a novel dataset collected in Experiment 2 was used as the new testing dataset to validate the out-of-sample predictive accuracy of different models. In addition to regularisation, regression techniques including subset selection and dimension reduction have also been used for robust regression of the high-dimensional dataset (a dataset in which the number of features is close to or larger than the number of observations). It will be tested whether other multivariate approaches would better fit the dataset in this study.

In this chapter, the different regression techniques were introduced in Section 6.2, and the analysis procedure and the criteria for comparison were described in Section 6.3. Regression techniques were compared in terms of predictive accuracy, model fit, and variable rankings and selections (Section 6.4, 6.5). The advantages and disadvantages of different methods were then discussed, to what extent facial attractiveness could be modelled from colorimetric facial traits was further explored, and future recommendations were given in Section 6.6.

6.2 Modelling techniques

The relationships between facial attractiveness and colour characteristics in the Chinese dataset were modelled using the eight statistical and machine learning algorithms as described below. The ordinary least squares regression was included for comparison. The other seven strategies proposed were based on the three most commonly used multivariate techniques, subset selection, dimension reduction, and regularisation. For all

the regression techniques that have tuning parameters, a ten-fold cross-validation was performed to determine the optimal parameters with the maximized model fit and optimise the algorithms. All the analyses were carried out in R (RDC, 2010).

Ordinary Least Squares Regression (OLS)   The OLS model was built using the *lm()* function under the *stats* library. In OLS, all the relevant colour predictors were included in one regression model without any process of variable selection.

6.2.1 Subset selection regression

In this study, both forward stepwise and backward stepwise methods were tested and the subset selection was achieved by an iterative procedure based on Akaike Information Criterion (AIC). The stepwise regression was implemented using the functions *ols_step_backward_aic()* and *ols_step_forward_aic()* in the *olsrr* package.

Stepwise Regression – Forward Steps (SF)   The model starts with only one intercept and then adds colour predictors based on AIC in a stepwise manner until AIC is no better.

Stepwise Regression - Backward Steps (SB)   The model starts with all the relevant colour predictors and then removes colour predictors based on AIC in a stepwise manner until AIC is no better.

6.2.2 Dimension reduction regression

The method doesn't perform feature selection but overcomes the problem of multi-collinearity by dimension reduction. PCR and PLSR are common dimension reduction regression techniques. In this study, both were conducted by the pls package using the function *pcr()* and *plsr()*, respectively.

Principal component regression (PCR) In this study, a ten-fold cross-validation was utilised to determine the number of principal components by minimizing the Root Mean Squared Error (RMSE) of the prediction on the one-fold new data.

Partial least squares regression (PLSR) The ten-fold cross-validation was also adopted to select the optimal number of linear combinations (latent components) for PLSR by minimizing the Root Mean Squared Error (RMSE) of the prediction on the one-fold new data.

6.2.3 Regularisation regression

Techniques of regularisation or penalisation are used to constrain the coefficients of a model and reduce the variance of the parameter estimators, which have become popular for the analysis of high-dimensional datasets in recent years. The ridge and lasso were done by the *glmnet* package and the elastic net was done by the *caret* package.

The Ridge Regression (RR)   RR applies the L2 penalty to minimise overfitting and reduces the coefficients of less important variables to approach 0. The higher value of the shrinkage parameter, lambda, the more aggressively the coefficients are shrunk toward zero. Here, the optimal lambda was defined using a ten-fold cross-validation process while alpha = 0.

Least Absolute Shrinkage and Selection Operator Regression (LASSO)   LASSO applies the L1 penalty to minimise overfitting and reduces the coefficients of less important variables to 0 (whereas RR make coefficients approach 0). The same ten-fold cross-validation process was performed to determine the tuning parameter, lambda, while alpha = 1.

The Elastic Net Regression (EN) EN combines the penalty terms of RR and LASSO. In EN, both parameters, alpha and lambda, can be tuned to optimize the model fit where alpha controls the degree to which the model shrinks coefficients and lambda determines how aggressively coefficients are set to zero. The ten-fold cross-validation was implemented to generate the best combination of alpha and lambda with the maximised fit (minimised RMSE).

6.3 Analysis procedure and criteria for model comparison

6.3.1 Analysis procedure

To provide a repeatable analytical tool for facial attractiveness modelling and conduct rational comparisons across different regression techniques, a four-step analysis procedure was followed, as shown in Figure 6.1.

Step 1: Variable Identification	• Literature Review • Facial Colour Analysis • N = 65
Step 2: Pre-processing	• Correlation Analysis • Irrelevant Variables Removed • N = 21
Step 3: Facial Attractiveness Modelling	• Ordinary Least Squares (OLS) • Subset Selection (SF, SB) • Dimension Reduction (PCR, PLSR) • Regularization (RR, LASSO, EN)
Step 4: Comparison of Regression Techniques	• Predictive Accuracy • Goodness of Fit • Selection of Colour Predictors

Figure 6.1 The framework of the analysis procedure. N is the number of colour predictors.

Taking the Chinese dataset as an example, the relationships between attractiveness and facial colour cues were studied in this chapter. The first two steps were the same for all the modelling methods, which were to involve all the possible explanatory variables of facial colour characteristics and to remove the irrelevant variables. To identify as many as possible colour variables in the first step, in addition to the colour cues mentioned in Section 5.2, this chapter also included all the local skin colours (as outlined in Figure 5.1-a), feature colours (as outlined in Figure 5.1-b), and the chroma, C^*, and hue angle, h_{ab} for all the colour variables as they may also be important colour parameters in relation to perception. A total of sixty-five explanatory colour variables were included in step 1. Then the correlation matrix was obtained to assess the association between attractiveness ratings and the sixty-five facial colour characteristics (Figure 6.2). The twenty-one colour variables that have significant correlations ($p < 0.05$) with attractiveness ratings were selected as valid predictors for the next step of modelling (marked as red in Figure 6.2). This step of data pre-processing was to remove the irrelevant variables from a large number of variables before further building the regression model as they might not make contributions to the prediction model.
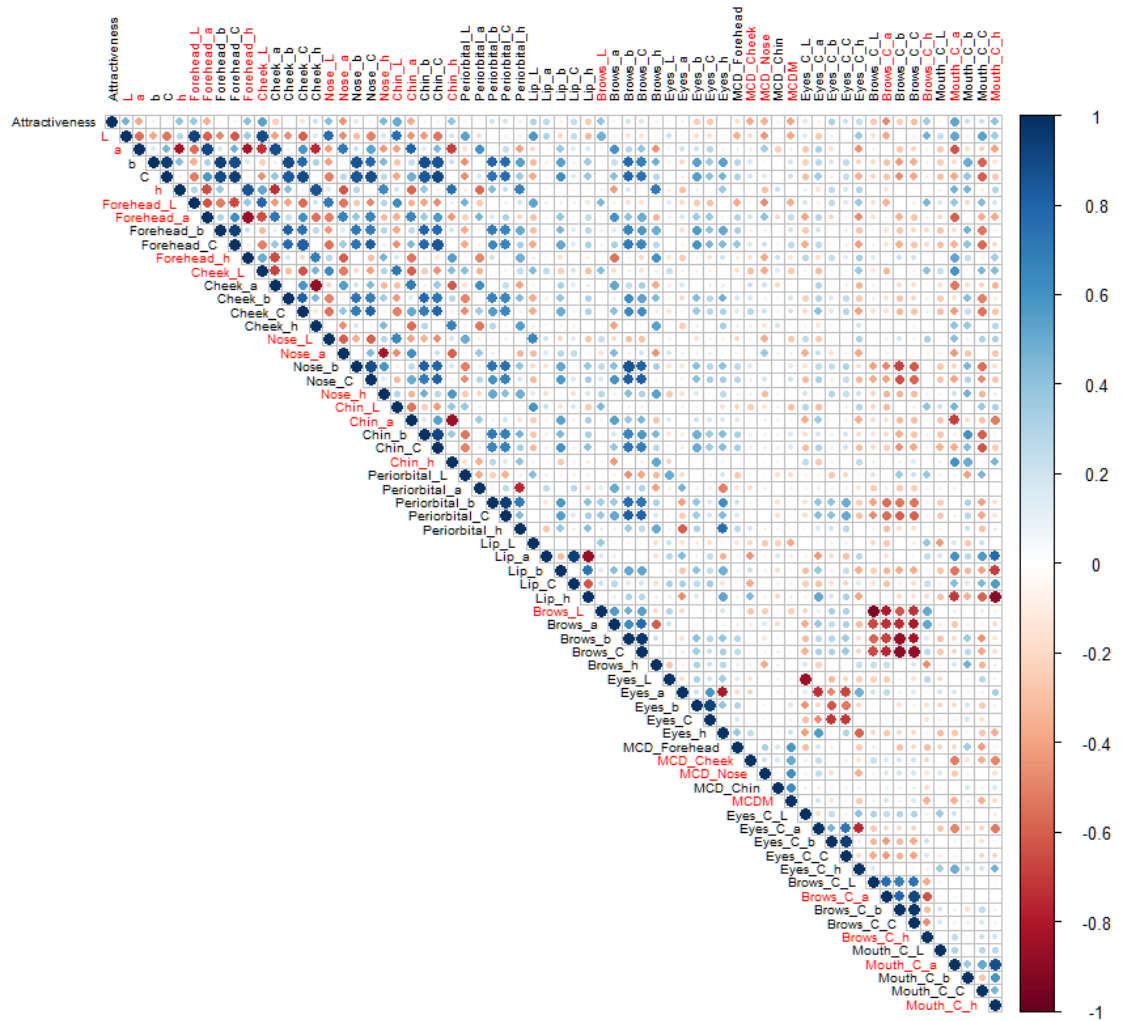
Figure 6.2 The correlation matrix between facial attractiveness ratings and facial colour characteristics. The twenty-one colour variables that have significant correlations (p < 0.05) with attractiveness ratings were marked in red.

The relationship between facial attractiveness and the relevant colour characteristics was then modelled in step 3 and model comparison was conducted in step 4. All the colour predictors were standardised to have zero mean and unit standard deviation before modelling. For both the training dataset (data collected in experiment 1 using the 40 Chinese images) and the testing dataset (data collected in experiment 2 using the new set of 60 Chinese images), ratings were averaged across all observers to create a score for each face before modelling from the face level colour traits.

6.3.2 Criteria for model performance comparison

Model performance was evaluated by the predictive accuracy, the goodness of fit, and the selection of colour predictors. To evaluate the effectiveness of different multivariate analysis techniques, after developing the eight models based on the training dataset, a novel testing dataset (using both new

observers and a new set of facial images) was used to test the performance of different models. Both the in-sample and the out-of-sample model performance were reported. On the other hand, the number of variables and the valid colour predictors that remained in each model were compared and discussed.

Predictive accuracy The Root Mean Square Error (RMSE) was adopted to measure the difference between the observed values (attractiveness scores rated by observers) and modelled values (attractiveness scores predicted by models) and compare the predictive accuracy of different models. by contrast with R^2, RMSE is not inflated by the number of predictors and it has the same unit as the original scale used in this study.

Goodness of fit   The model accuracy was also measured using the coefficients of determination R^2. It was calculated as the square of the Person correlation between the observed values (attractiveness scores rated by observers) and modelled values (attractiveness scores predicted by models).   The R^2 values were calculated for the training dataset, as the goodness of fit, and also for the testing dataset, as the goodness of prediction.

Selection or ranking of colour predictors For the regression techniques that perform direct variable selection such as SF, SB, LASSO, and EN, the model was evaluated by the number and selection of colour predictors that remained in the model. For the rest methods including OLS, PCR, PLSR, and RR, all the variables remained in the model and the rank of the variables was compared according to the standardised regression coefficients.

6.4 Predictive accuracy and model fit

The results of the in-sample and the out-of-sample model performance of the eight regression methods were shown in Table 6.1. Differences in RMSE values between different models were relatively smaller for the training dataset (from 0.42 to 0.62) but larger for the testing dataset (from 0.66 to 1.35). The range of R^2 for the training dataset across different models was from 42.6% to 73.9% and is always lower on the test as expected. In this study, PCR selected two principal components for optimal model fit, and the two components explained 58.6% of the variance in the original predictors and 42.6% of the variance in attractiveness; PLSR selected only one component, which explained 41.5% of the variance in the original predictors and 44.8% of the variance in attractiveness. Within the training dataset, OLS

and subset selection regressions had lower RMSE values and higher R^2 values than dimension reduction and Regularisation techniques. However, the Regularisation techniques achieved better out-of-sample model performance than all the other models, where EN showed the lowest RMSE value and RR had the highest R^2 value for the testing dataset. Although OLS showed the lowest RMSE and highest R^2 value for the training dataset, it performed the worst on the testing dataset.

Table 6.1 Comparison of the eight multivariate regression algorithms based on the RMSE and R^2 for the training dataset and the testing dataset.

Algorithms	Training RMSE	Training R^2 (%)	Testing RMSE	Testing R^2 (%)
OLS	0.42	73.9	1.35	10.8
Subset selection				
SF	0.51	61.2	0.82	38.5
SB	0.44	71.4	1.17	12.0
Dimension reduction				
PCR	0.62	42.6	0.71	39.9
PLSR	0.61	44.8	0.68	39.6
Regularisation				
RR	0.60	51.8	0.67	43.5
LASSO	0.54	58.1	0.70	39.4
EN	0.55	56.9	0.66	41.8

The accuracy of the eight regression models in predicting facial attractiveness was also demonstrated in the bar plots in Figure 6.3. The OLS and subset selection regressions showed lower in-sample RMSE values but higher out-of-sample RMSE values. The models using dimension reduction and regularisation techniques were just the opposite, resulting in closer RMSE values between the training dataset and the testing dataset.
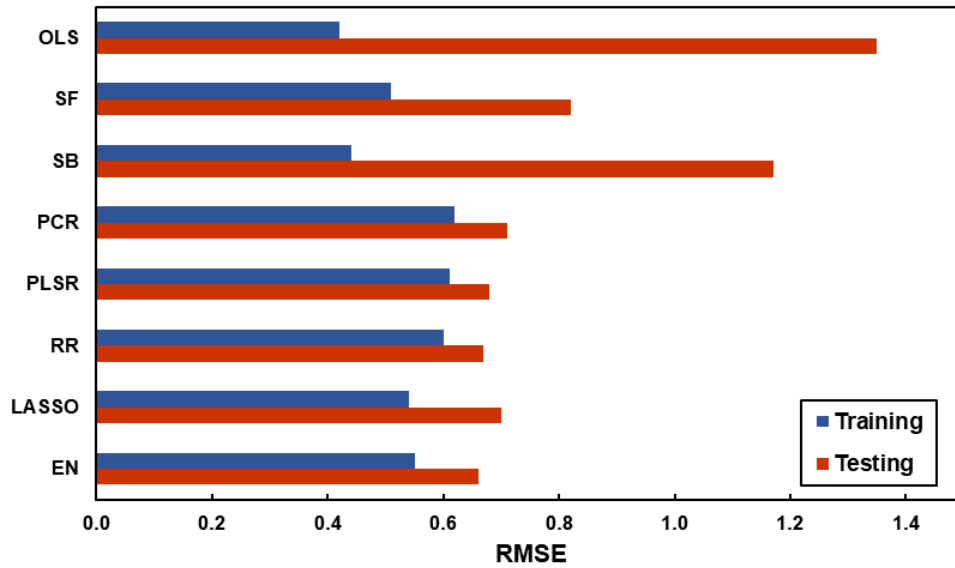
Figure 6.3 The RMSE values of the eight regression models in predicting facial attractiveness for the training data (blue bars) and the testing dataset (red bars).

The scatter plots in the next two pages gave the comparisons between the actual values of facial attractiveness ratings recorded during the experiments and the predicted values of facial attractiveness calculated from the regression models of all the facial images. Each red data point on the left column represented one of the forty facial images that were judged in Experiment 1, and each blue data point on the right column represented one of the sixty facial images that were judged in Experiment 2. The lower level of dispersion of the data points from the diagonal lines suggested the better performance of the model. Generally, different models did not show considerable differences in the attractiveness prediction of the training dataset, whereas different degrees of dispersion were shown across different regression models indicating the different out-of-sample model performances as mentioned above.
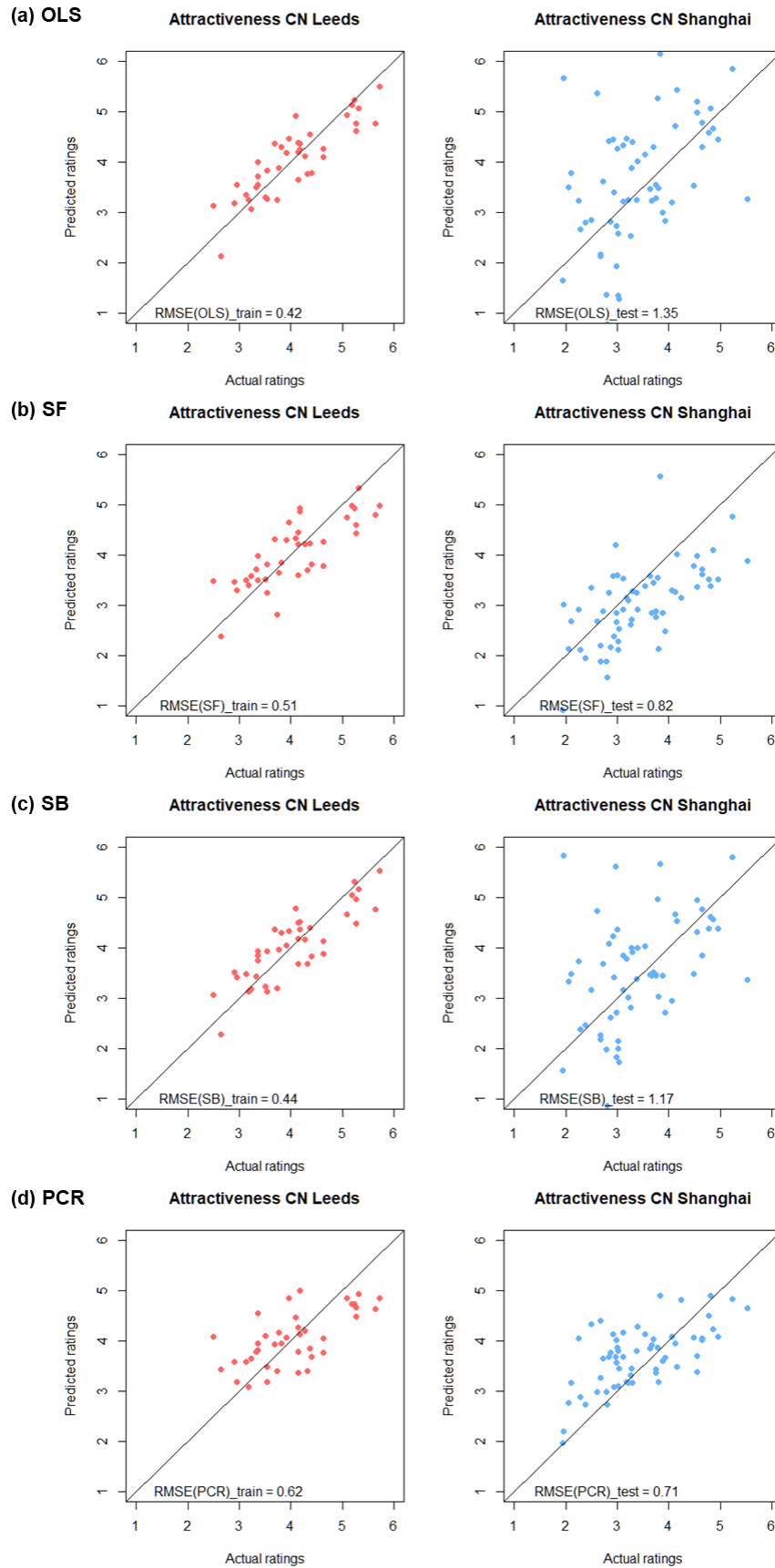
Figure 6.4(a)-(d) Model performance of the (a) OLS, (b) SF, (c) SB, (d) PCR in predicting facial attractiveness for the training data (left column) and the testing data (right column).
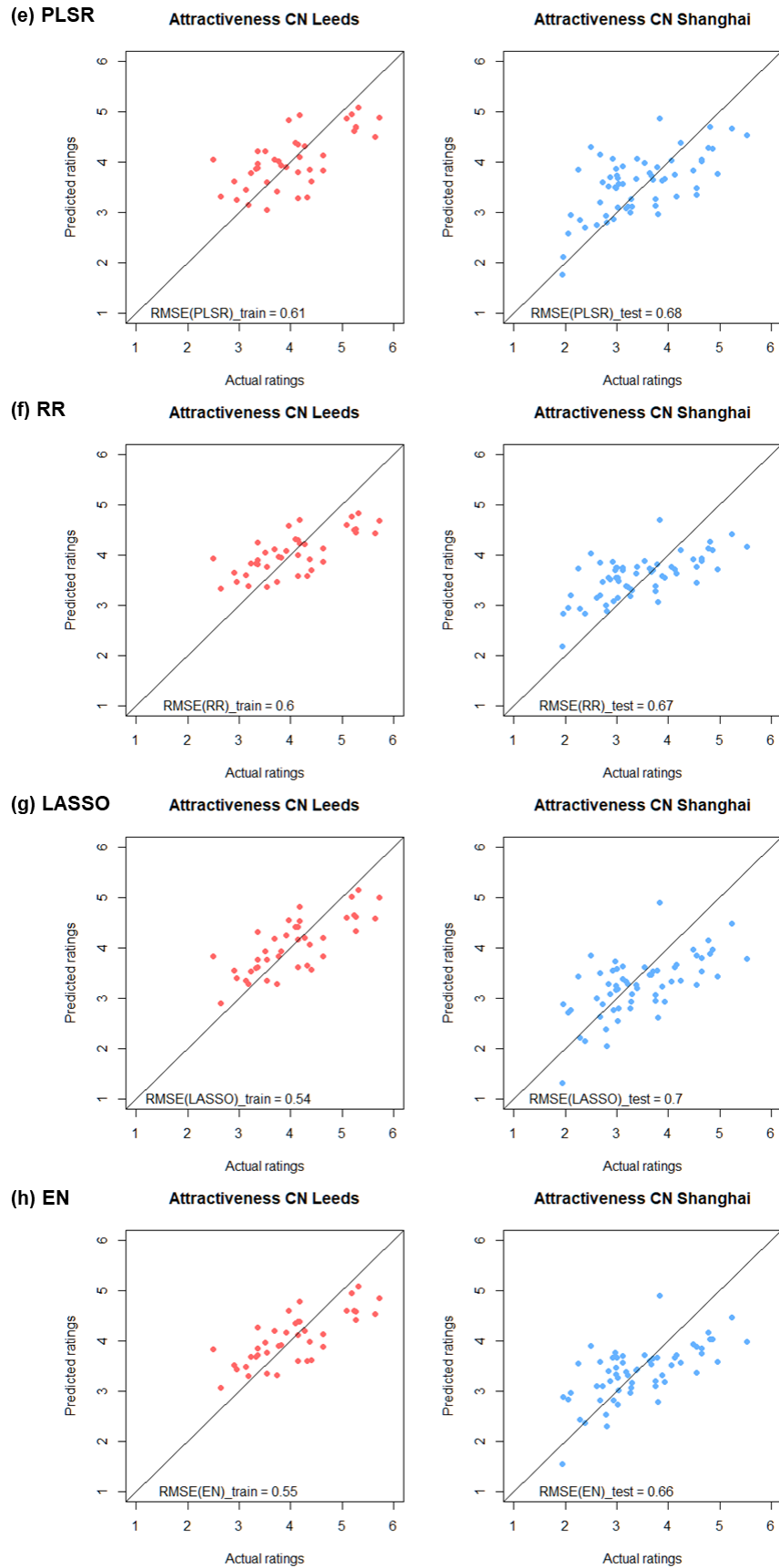
Figure 6.4(e)-(h) Model performance of the (e) PLSR, (f) RR, (g) LASSO, (h) EN in predicting facial attractiveness for the training data (left column) and the testing data (right column).

6.5 Ranking and selection of predictors

The numbers of colour predictors that were selected by SF, SB, LASSO, and EN models were 7, 11, 11, and 14, respectively. For OLS, PCR, PLSR, and RR models, all 21 variables remained in the model. Based on the standardised regression coefficients in all regression models, the colour predictors were ranked in each model. Considering the regularisation techniques gave the better predictive accuracy as described in the previous section, the top 11 colour predictors selected by both LASSO and EN were listed in Table 6.2, ordered by the relative importance of each colour predictor (according to absolute standardised regression coefficients) in the EN model (last column). The rankings of these 11 colour predictors in eight regression models were given in the table. Some variables that were not selected by SF or SB were marked as 0. Between RR, LASSO, and EN, almost all the top 11 variables selected were in common.

Table 6.2 Ranking of the eleven colour predictors selected by LASSO & EN in the eight regression models. Variables that were not selected by SB or SF regression are marked as 0.

LASSO & EN selected	OLS	SF	SB	PCR	PLSR	RR	LASSO	EN
Brows_C_a*	5	4	3	8	5	1	1	1
MCDM	12	6	4	9	11	3	2	2
Nose_h_{ab}	7	0	6	18	7	2	6	3
MCDM_Cheek	10	7	0	15	17	4	7	4
Mouth_C_a*	17	0	7	10	2	5	8	5
Chin_L*	13	10	0	4	3	6	9	6
Forehead_h_{ab}	4	5	0	19	9	11	5	7
Chin_h_{ab}	8	0	0	21	14	10	3	8
Cheek_L*	2	2	2	3	1	7	11	9
Nose_L*	16	8	5	6	8	12	4	10
Mouth_C_h_{ab}	15	9	0	16	10	9	10	11

The correlation matrix between the top eleven colour predictors selected by LASSO & EN and facial attractiveness ratings were further visualised in the heatmap in Figure 6.5. To identify the hidden structure and pattern in the matrix, the 11 colour predictors were reordered based on the hierarchical clustering as shown in the five black boxes, which were linked to the brows colour contrast (a*), skin colour variation (overall or cheek), local skin hue angle (forehead, chin or nose), the mouth colour contrast (a*) and local skin lightness L* (nose, chin, or cheek), from top left to bottom right.
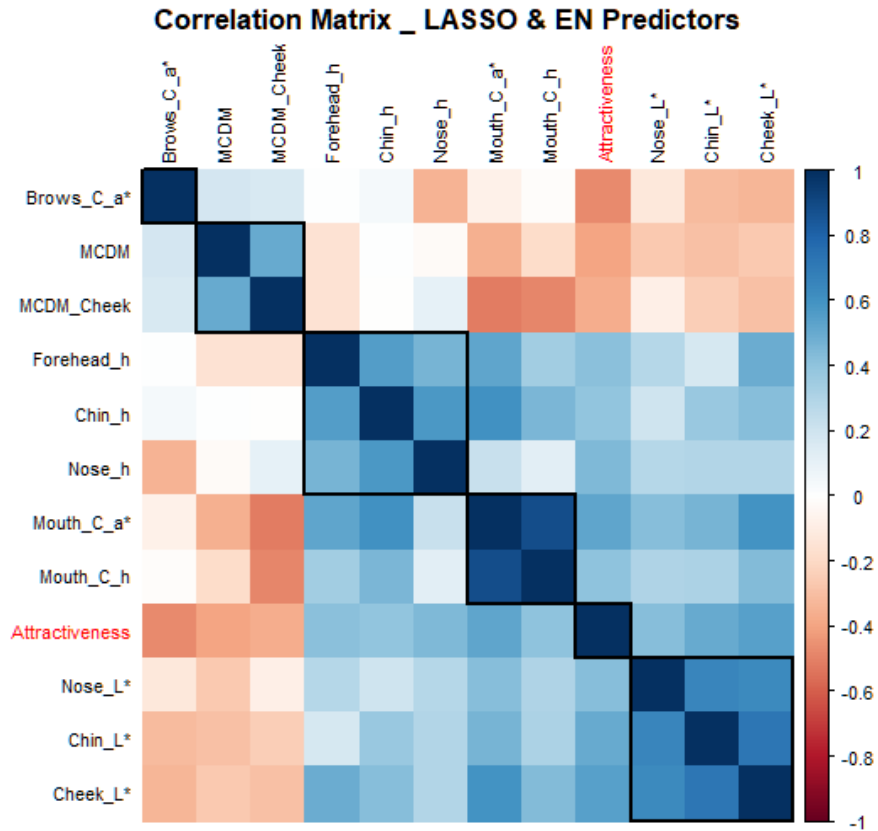
Figure 6.5 The correlation matrix between the eleven colour predictors selected by LASSO & EN and facial attractiveness ratings.

6.6 Discussion

6.6.1 The analytical framework for modelling facial attractiveness from various facial colour cues

Previous chapters demonstrated that using colorimetric cues in isolation to study facial attractiveness was associated with various problems, especially simplifying the complex nature of facial attractiveness judgement based on various colour cues (Section 5.6, 5.7). Therefore, the univariate approaches that were almost exclusively employed in the past could not satisfy the current research purpose of attractiveness modelling in real faces. To consider all the relevant colour characteristics simultaneously and manage the complex data structure with a large number of correlated colour features, multivariate approaches are needed. An important challenge is to develop an analytical tool that can be used to select valid predictors and make accurate predictions. The last chapter was an attempt to involve a relatively large number of colour variables and predict preference by the EN model. Due to the high number of candidate variables and their correlations, predicting attractiveness from various colour cues could be difficult and it

was unknown whether other multivariate approaches would better fit the dataset.

In this chapter, more potential colour predictors were included which has never been considered together before. Different multivariate statistical techniques were compared in terms of the effectiveness and performance in attractiveness modelling. A complete analytical framework was provided for modelling facial attractiveness from a large number of facial colour characteristics (see Figure 6.1). Starting from identifying relevant variables, all the potential colour predictors were summarized based on a literature review and a pre-processing step of correlation analysis was carried out to ensure all the colour characteristics contributing to attractiveness judgement were included in the following regression analysis. Different regression techniques were then employed and the models were estimated using a training dataset and then validated using a novel testing dataset.

This methodology can be applied in any future research on attractiveness modelling from various facial colour traits. If any other facial colour characteristic is identified to have an impact on facial attractiveness judgment, the new colour predictors can be added to the analysis from the first step of variable identification. Any other potential regression techniques can also be considered in the third step of modelling to see whether the final performance of the model will be improved. Besides, following this framework, other facial preference attributes can be studied as well and the prediction model can be built considering different ethnic groups, ages, genders, etc.

6.6.2 Comparison of different regression techniques

Different regression models were built to predict preference ratings from a large set of colour variables. Due to the collinearities between those explanatory variables (see Figure 6.2), conventional regression methods such as OLS may cause problems of multi-collinearity and result in overfitting. The problem of multi-collinearity was addressed using three different approaches, reducing the number of independent variables, removing the dependency between variables by projecting them into a new uncorrelated variable space, and adding some regularisations. To avoid the issue of overfitting, a novel testing dataset was used and any overfitting caused by the high-parameter model would only help fit the training dataset, but not the testing dataset.

The stepwise multiple regression models directly used the colour predictors as independent variables for facial preference prediction. The advantage of such an approach is that the model is straightforward and easy to run and interpret. However, among the three regression techniques, stepwise multiple regression was most largely affected by multi-collinearity and overfitting. Especially the backward method, as it contained more variables than the forward method, the model resulted in an out-of-sample predictive error almost three times larger than the in-sample error (RMSE values). The disadvantage of the subset selection methods is notable that they are not designed to handle high dimensional datasets and can easily fail to predict the attractiveness when the number of predictors increases. The methods assume that there is a perfect model, but due to the correlations between independent variables, the results vary based on the order of the variable selection which made it difficult to identify the most important contributor in the model.

Dimension reduction methods overcome the problem of multicollinearity by transforming the original predictors into uncorrelated principal components. As a result, the out-of-sample model accuracy of both PCR and PLSR was relatively closer to the in-sample model accuracy compared to other methods. Meanwhile, as the PCR and PLSR models selected the most important principal components as the independent variables (in this study, PCR selected the first two principal components and PLSR only selected the first component), the model had the least number of predictors and the least degrees of freedom. The shortcoming of selecting principal components is that a certain amount of colour information has been lost during the dimension reduction process. Such loss of information influences the R^2 values so that both the PCR and PLSR models explained the lowest variance in attractiveness (42.6% and 44.8%, respectively) than all the other methods. Regression models using dimension reduction techniques also have the disadvantage that they are difficult to interpret and are not practical to make future predictions. Though the model only contains a small number of principal components, any future prediction will require the analysis of all the relevant colour variables to calculate those principal components.

The regularisation regression models are easier to understand and interpret. Especially the LASSO and EN, as they performed feature selection, once the model has been developed only a subset of facial colour characteristics with the strongest effects on the response variable will need to be analysed for future facial preference prediction and will be directly used to build the

model. Both models gave relatively accurate predictions on the testing dataset while EN performed better in terms of both predictive error (RMSE) and variance explained (R^2). RR also gave a similar performance as EN but with all the colour predictors remaining in the RR model. To make future predictions, the model will need all the relevant colour predictors of any new faces to be calculated. Generally speaking, based on both the predictive accuracy and the model fit, the regularisation regressions could be the best-fit technique for modelling attractiveness from facial colour traits.

6.6.3 The variable rankings and selections across regression models

Variable rankings and selections are different issues from the model performance which need to be simultaneously taken into account. According to the complete correlation matrix between facial attractiveness ratings and the facial colour characteristics (Figure 6.2), all the relevant colour predictors at significant levels were L*, a*, or h_{ab} related without an exception. These results revealed that Chinese observers relied more on colour cues related to skin lightness (L*), redness (a*) or hue angle (hab) for attractiveness judgement. The other two colour attributes, yellowness (b*) or chroma (C*) in CIELAB colour space, are less important and almost entirely unused to make decisions on attractiveness.

Among these relevant colour predictors, the variable ranking of the RR, LASSO, EN, and PLSR showed a large overlap. LASSO and EN selected the same top eleven colour predictors. Within these 11 colour predictors, the top 10 from RR, the top 9 from PLSR, the top 6 from PCR, the top 6 from SF, and the top 8 from SB were included (Table 6.2). After grouping the correlated variables (Figure 6.5), the most important features were clustered into five groups: the brows colour contrast (a*), skin colour variation (overall or cheek), local skin hue angle (forehead, chin or nose), the mouth colour contrast (a*) and local skin lightness L* (nose, chin, or cheek).

Generally, the results were consistent with the Chinese model of facial attractiveness in the last chapter (Section 5.6) except for the two new predictors, local skin hue angle and local skin lightness. Both the hue angle, h_{ab} and the local skin colour attributes were newly added to the analysis of this chapter as potentially important colour parameters in relation to perception. The results verified the assumption that h_{ab} is an important colour predictor of attractiveness evaluation. And calculating the skin lightness of some local skin areas might be enough for future attractiveness predictions instead of analysing the overall facial lightness.

6.6.4 The predictive accuracy in attractiveness modelling from colorimetric facial traits

Comparing the model performance, the research question 'how accurate facial attractiveness can be predicted from colorimetric facial traits' was addressed. In this study, different regression techniques were able to predict facial attractiveness with a minimum predictive error of less than one point (0.66 points) on a 7-point rating scale (EN) and a highest prediction R^2 of 43.5% (RR) both when validated on an independent dataset. The RMSE was slightly smaller than the Chinese model of attractiveness in the last chapter (RMSE=0.71, see Section 5.6) and the possible reason could be that more variables were involved and identified as important colour predictors in this chapter (e.g. the hue angle, h_{ab}) and thus the model performance was improved.

As discussed in Section 5.7.4, studies that used univariate models to predict attractiveness from single colour cues may underestimate the role of facial colour cues in the attractiveness judgement of real faces (Foo, Rhodes, et al., 2017; Jones, 2018; Tan et al., 2018; Cai et al., 2019). The current multivariate approaches again showed superior performance and confirmed the effectiveness of colour in attractiveness modelling. On the other hand, very few studies used multivariate approaches to build facial attractiveness models based on structural facial features including averageness, dimorphism, and symmetry, and their out-of-sample RMSE varied from 0.46 to 0.77 (Jones and Jaeger, 2019; Holzleitner et al., 2019). Compared to those studies, the colour-based models in this study showed comparable importance of the colorimetric facial traits in attractiveness judgement.

6.6.5 Future recommendations

An ideal regression model for attractiveness prediction should be sparse, interpretable, and well-predictive. Based on the discussion above, both regression techniques of dimension reduction and regularisation can be used for future study of attractiveness modelling from facial colour traits. Determining the specific regression algorithm depends on the investigatory priority. The PCR, PLSR, and RR have been identified as suitable algorithms to deal with multi-collinearity (Geladi and Kowalski, 1986; Abdi and Williams, 2010). The LASSO and EN select fewer variables and are easier to interpret and more practical to implement. Complex models are not necessarily performing better than simpler ones (Ransom et al., 2019). For evaluating the overall performance of different algorithms in the future, both the

simplicity of the model and the predictive accuracy need to be taken into account.

6.7 Summary

This chapter focused on the comparison of different regression techniques for attractiveness modelling using a large number of facial colour cues. A summary of the analysis and major findings are given below:

- An analytical framework was proposed, following which eight regression models of three regression techniques were compared in terms of predictive accuracy, the goodness of fit, and variable selection.

- The framework of analysis procedure could serve as an important analytical tool for future study in facial attractiveness or other preference attributes modelling and prediction from a large number of facial colour characteristics.

- Generally, both dimension reduction regression and regularisation regression outperformed the classical OLS and stepwise techniques as the high variance and overfitting weakened the prediction power of these classical approaches.

- The EN showed the most accurate predictions when validated using the independent testing dataset. With fewer colour variables selected, the model is also easier to interpret and more practical to implement.

- The most important colour features of attractiveness identified in this study were: the brows colour contrast (a*), skin colour variation (overall or cheek), local skin hue angle (forehead, chin or nose), the mouth colour contrast (a*) and local skin lightness L* (nose, chin, or cheek).

Chapter 7 Overall facial colour appearance

7.1 Overview

The last three chapters have revealed that facial colour appearance could largely influence the facial attractiveness assessment. Previous research on facial preference as well as this study has commonly used the average facial skin colour specified in CIELAB colour space (L*, a*, and b*) to roughly represent the overall facial colour appearance, which was based on the colorimetric averaging hypothesis. However, it was unknown whether the colorimetric average could represent the perceptual overall colour appearance of human faces, and whether the CIELAB coordinates were suitable for quantifying the colour perception of human skin. In the next two chapters, efforts were made to better understand the perception of facial colour appearance.

This chapter addressed the research question 'what is the overall colour appearance or the globally representative colour of a human face'. Using the colour appearance matching data collected in Experiment 3, statistical analysis (Section 7.2) was done to test the colorimetric averaging hypothesis (Section 7.3). Mathematical models were built to accurately predict the perceptual colour appearance of human faces (Section 7.4). The influential factors of skin appearance perception and the perceptual difference between observers were discussed (Section 7.5).

7.2 Statistical analysis

In Experiment 3, 40 Caucasian and 40 Chinese facial images were used. The average pixel colour of each facial image was obtained by image analysis previously (see Section 4.2).

During the experiment, the overall colour appearance of each facial image was assessed by colour appearance matching using uniform face-shaped colour patch by a panel of observers. The matched patch colour of each image from each observer was recorded in terms of display RGB values. The mean matched colour over all observers for each image was then calculated (by face dataset). The RGB values were transformed into the CIE XYZ tristimulus values based on the display characterisation model (the forward model) and subsequently transformed to CIELAB colour coordinates using display white point as the reference white.

The average colour and matched colour were first compared, and their colour appearance difference was calculated in CIELAB uniform colour

space for each face. The one-sample t-test (H_0: no colour change) was used to evaluate the overall colour changes using the by face dataset. To further test the effects of the face ethnicity, observer ethnicity, and the interaction between the two on the colour change applied, the linear mixed-effect model was set up using the raw data collected (dependent variables = colour change in three dimensions, $\triangle L^*$, $\triangle a^*$, $\triangle b^*$; fixed effects = face ethnicity, observer ethnicity; random effects: face ID, observer ID). The main effects and the interaction between face and observer ethnicity were included in the model. At last, based on the consistent colour shift from the average pixel colour and the matched facial colour across different faces, simple linear regressions were performed to quantify the overall facial colour appearance.

All the analyses were conducted in R (RDC, 2010). The one-sample t-test was done using the *t.test()* function in the *stats* R package. The linear mixed-effect models were implemented by the *lmer()* function in the *lme4* R package (Bates et al., 2015). The simple linear regression model was built using the *lm()* function under the *stats* library.

7.3 Average colour vs. matched colour

The mean colour difference (with the standard deviation) between the average colour and the matched colour of 80 facial images is 3.14 ± 0.56 ΔE^*ab unit (Caucasian faces: 3.12 ± 0.52 ΔE^*ab; Chinese faces: 3.17 ± 0.58 ΔE^*ab). Table 7.1 shows the results of the one sample t-test and the mean colour changes in terms of the five individual colour attributes (L^*, a^*, b^*, C^*_{ab} and h_{ab}) and the overall colour difference. One sample t-test revealed that the patch colour significantly changed after matching. Observers increased facial lightness (L^*) by 3.00 units (SE = 0.06), $t(79) = 46.965$, $p < 0.001$, decreased facial redness (a^*) by 0.77 units (SE = 0.04), $t(79) = -21.133$, $p < 0.001$, and decreased facial yellowness (b^*) by 0.27 units (SE = 0.03), $t(79) = -8.907$, $p < 0.001$, to match the overall colour appearance of the faces based on their perceptions. In other words, observers decreased chroma (C^*_{ab}) by 0.56 units (SE = 0.03), $t(79) = -18.949$, $p < 0.001$, and increased the hue angle (h_{ab}) by 1.69 units (SE = 0.06), $t(79) = 51.094$, $p < 0.001$, to match the overall colour appearance of the faces based on their perceptions.

Table 7.1 Summarized statistics from one sample t-test: overall colour change after matching.

Colour Axis	Significance	Mean ± SE
L*	t_{79} = 46.965; p<0.001***	3.00 ± 0.06
a*	t_{79} = -21.133; p<0.001***	-0.77 ± 0.04
b*	t_{79} = -8.907; p<0.001***	-0.27 ± 0.03
C^*_{ab}	t_{79} = -18.949; p<0.001***	-0.56 ± 0.03
h_{ab}	t_{79} = 14.886; p<0.001***	1.69 ± 0.11
Overall △E	t_{79} = 51.094; p<0.001***	3.14 ± 0.06

*P≤0.05, ** P≤0.01, ***P≤0.001.

Figure 7.1 visualises the colour shift from the average pixel colour (hollow points) to the matched facial colour appearance (solid points) of 40 Caucasian faces (blue lines) and 40 Chinese images (orange lines) in a*b* plane and L*C* plane in CIELAB uniform colour space. The actual perceived facial colour appearance by observers was obviously different from the average facial colour. The colour shift of both Caucasian faces and Chinese faces showed a consistent trend that the actual matched facial colour appearance had lower a* values, slightly lower b* values and much higher L* values compared to the average facial pixel colour.
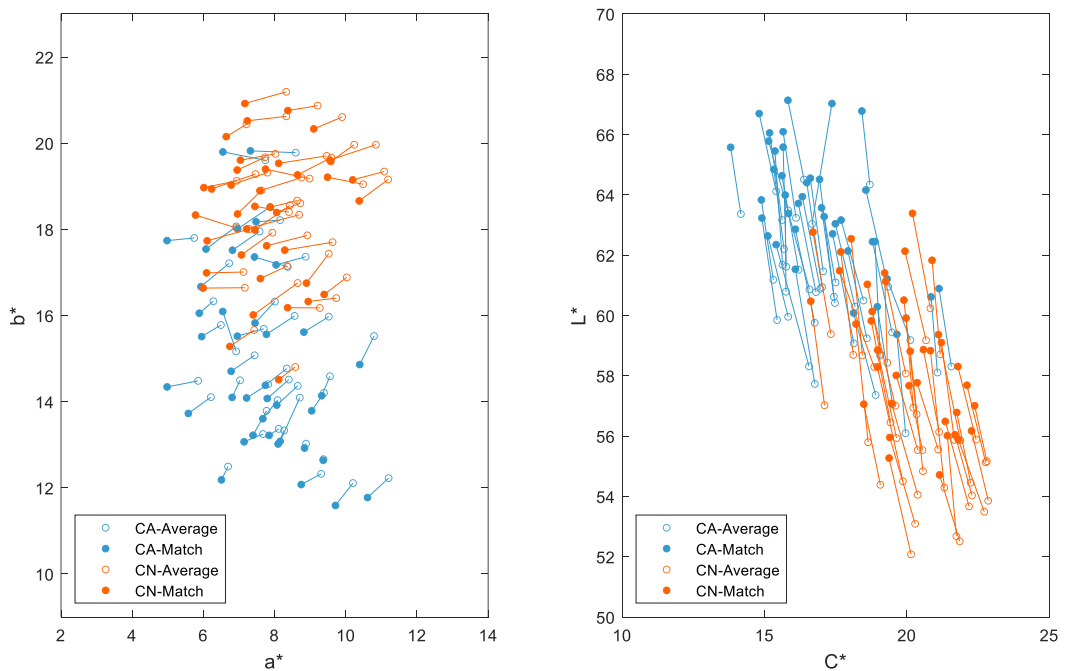
Figure 7.1 Colour shift from the average colour (hollow points) to the matched colour (solid points) of Caucasian faces (blue lines) and Chinese faces (orange lines) in CIELAB a*b* space (left) and L*C* space (right).

To further explore how the colour change applied was influenced by the face ethnicity or observer ethnicity, linear mixed-effect modelling was carried out for the colour change in a*, b*, and L*, respectively. Figure 7.2 shows the difference in colour change between the face and observer of the two ethnic groups. After the colour matching, $\triangle$a*, $\triangle$b*, and $\triangle$L* were always negative, negative, and positive, respectively, across the four subgroups (Figure 7.2), showing the consistent pattern of colour change from the average pixel colour to the matched colour.

For the redness change ($\triangle$a*), there were main effects of face ethnicity ($F_{1,79}=77.807$; $p<0.001$) and observer ethnicity ($F_{1,42}=10.708$; $p<0.01$), but their interaction was not significant ($F_{1,3239}=0.122$; $p=0.727$). The significant main effects showed that both observers reduced redness more on Chinese faces than Caucasian faces to match the perceived facial colour appearance, and Chinese observers made a larger decrease than Caucasian observers (Figure 7.2a). For the yellowness change ($\triangle$b*), the face ethnicity was not significant ($F_{1,46}=0.002$; $p=0.963$), but the observer ethnicity ($F_{1,42}=13.479$; $p<0.01$) and the interaction ($F_{1,3239}=13.979$; $p<0.001$) was significant. Chinese observers tended to decrease more yellowness on the average colour to match the actual appearance colour, and both observers decreased yellowness more when they viewed faces of their own ethnicity (Figure 7.2b). For the lightness change ($\triangle$L*), the face ethnicity was also not significant ($F_{1,79}=0.427$; $p=0.515$), but the observer ethnicity ($F_{1,42}=13.247$; $p<0.01$) and the interaction ($F_{1,3239}=18.143$; $p<0.001$) was significant. Caucasian observers made a larger increase in lightness than Chinese observers on both faces, and both observers increased lightness more when they viewed faces of their own ethnicity (Figure 7.2c).
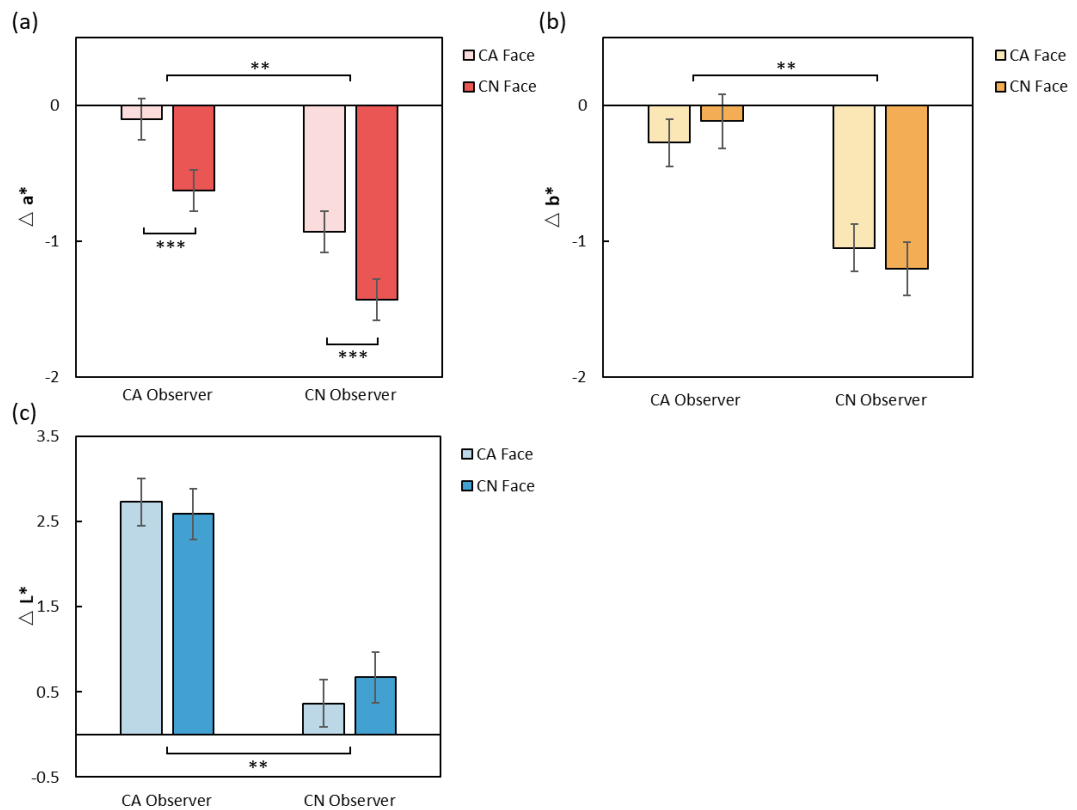
Figure 7.2 Bar plots showing the effects of face ethnicity, and observer ethnicity on the colour change applied on a* (a), b* (b), and L* (c). The error bars indicate 95% confidence intervals. *P≤0.05, ** P≤0.01, ***P ≤0.001.

Overall, the actual facial colour appearance perceived by observers was less reddish, slightly less yellowish, and much lighter compared to the average facial colour which was commonly used in previous research. The effect of observer ethnicity on facial colour appearance match was observed. Comparing the two groups of observers, the facial appearance perceived by the Chinese was less yellowish and less reddish than what was matched by Caucasian observers, while the appearance perceived by Caucasian observers was lighter than what Chinese observers matched.

7.4 Model the overall facial colour appearance

Based on the consistent colour change between the average pixel colour and the matched colour (Figure 7.1), it was possible to carry out a simple regression analysis to predict the actual facial colour appearance perceived by observers from the average pixel colour. Based on the linear mixed-effect analysis, the effect of face ethnicity was only significant when matching the facial redness (a*), thus separate linear regression models were set up to predict the facial redness (a*) of Caucasian faces and Chinese faces,

respectively. As for facial yellowness (b*) and lightness (L*), the same model was applied to the faces of both ethnicities.

Table 7.2 Linear regression model predicting the matched facial colour appearance from the average pixel colour.

	β	SE	t	P
Regression 1: a* (CA Face)				
Model	$F_{1,38}$ = 858.7, P <0.001***, Adjusted R^2=0.958			
(intercept)	-0.894	0.290	-3.077	0.004**
$\overline{a^*}$	1.039	0.035	29.304	<0.001***
Regression 2: a* (CN Face)				
Model	$F_{1,38}$ = 869.9, P <0.001***, Adjusted R^2=0.958			
(intercept)	-1.053	0.301	-3.495	0.001**
$\overline{a^*}$	1.011	0.034	29.495	<0.001***
Regression 3: b* (Both Faces)				
Model	$F_{1,78}$ = 6918, P <0.001***, Adjusted R^2=0.989			
(intercept)	-0.581	0.209	-2.777	0.007**
$\overline{b^*}$	1.019	0.012	83.172	<0.001***
Regression 4: L* (Both Faces)				
Model	$F_{1,78}$ = 2373, P <0.001***, Adjusted R^2=0.968			
(intercept)	2.811	1.202	2.339	0.022*
$\overline{L^*}$	1.003	0.021	48.712	<0.001***

Table 7.2 summarises the simple linear regression analysis for the perceived facial colour appearance, a*, b*, and L*, respectively. All the regression models were statistically significant at less than 0.001 level. The average pixel colour $\overline{a^*}$, $\overline{b^*}$, and $\overline{L^*}$ explained 95.8 per cent, 98.9 per cent, and 96.8 per cent of the variance in the prediction of the perceived facial colour appearance a*, b*, and L*, respectively. Figure 7.3 shows the linear regressions of the average pixel colour on matched colour appearance with regression lines. Each data point represents one face that was used in the experiment. Basically, all the data are clustered tightly around the regression line showing a strong linear relationship.
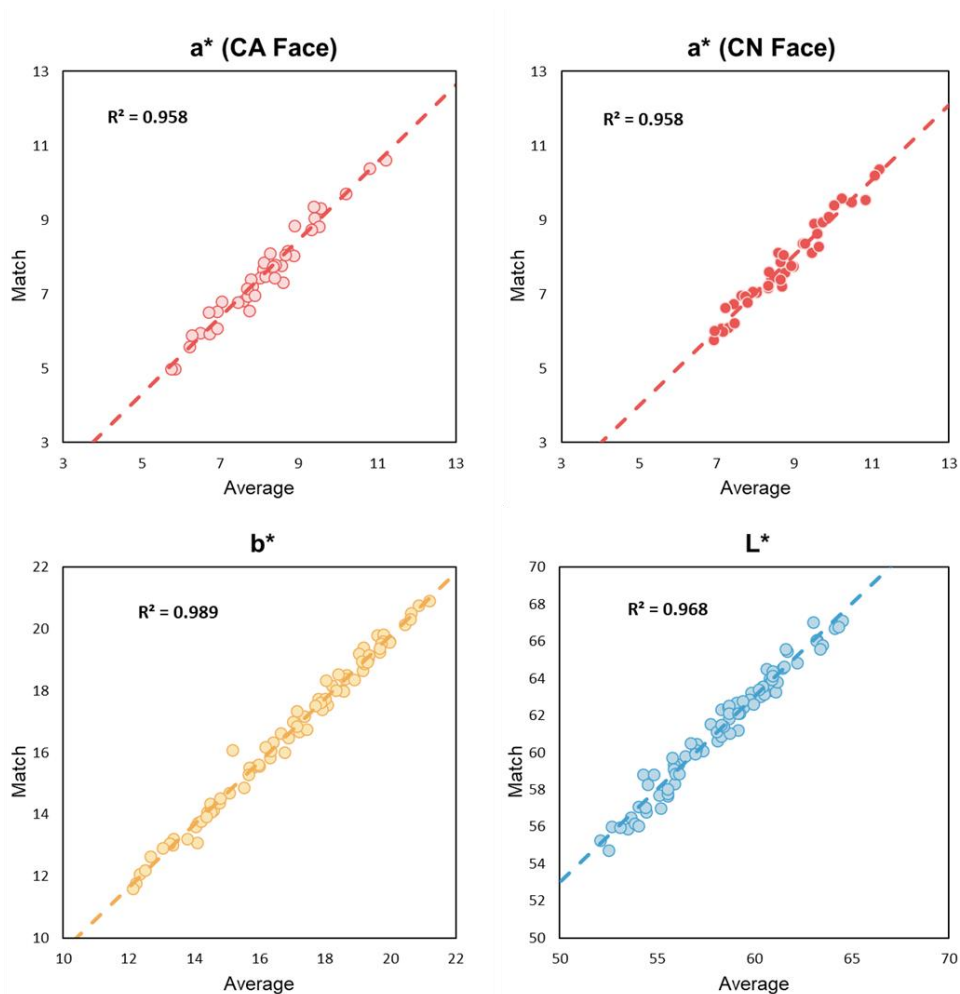
Figure 7.3 The linear regressions of the average pixel colour on the matched colour appearance with regression lines: a* of CA faces (upper left), a* of CN faces (upper right), b* (bottom left), L* (bottom right).

7.5 Discussion

7.5.1 The overall colour appearance of human faces

This study focused on the overall facial colour appearance with two research questions: (1) What is the perceived overall colour appearance of a human face? (2) How to quantitatively model this facial colour appearance? A colour matching experiment was conducted in this study and 80 real facial images were used as representative of the complexion and skin colour variation of real, Caucasian and Chinese, human faces. The main finding was that the overall facial colour appearance was significantly different from the average colour in the facial skin area though the latter was exclusively considered as the overall facial colour appearance in previous studies. Interestingly, the actual facial colour appearance perceived by observers was less reddish, slightly less yellowish, and much lighter compared to the average facial

colour. Furthermore, the actual facial colour appearance was accurately quantified and modelled for the first time in this study. The overall facial colour appearance a*, b*, and L* was defined can be predicted by simple linear regressions of average pixel colour $\overline{a*}$, $\overline{b*}$, and $\overline{L*}$ calculated in the facial skin area using the following equations:

$$a^* = 1.039\overline{a^*} - 0.894, \text{ for CA faces;}$$

$$a^* = 1.011\overline{a^*} - 1.053, \text{ for CN faces;}$$

$$b^* = 1.019\overline{b^*} - 0.581;$$

$$L^* = 1.003\overline{L^*} + 2.811.$$

Equation 7.1

The current models promised a high accuracy for facial colour appearance quantification with more than 95 per cent of the variance explained. These models can be used in the future prediction of the overall facial colour appearance of any new facial images. As the overall facial colour appearance describes what the colour stimuli of facial skin look like in human colour vision or the representative colour of a human face, the model can be applied in many applications that need to quantify or reproduce the colour appearance. For example, it can be used as an image-based method in facial colour measurement to evaluate dermatology treatment, skin care products, or personalised cosmetic shopping, replacing the conventional visual assessment methods of using skin colour charts or palettes (Fitzpatrick, 1988; Taylor et al., 2005; De Rigal et al., 2007; Swiatoniowski et al., 2013).

7.5.2 Factors that influence the perception of overall facial colour appearance

Seemingly, the current results didn't support the colorimetric averaging hypothesis which claims that the global colour impression is determined by the colorimetric average of the elemental colours in the multi-coloured textured patterns (Sunaga and Yamashita, 2007). If it's not the case, how did the overall colour perception evoke when the observers were presented with a human face and what were the factors that might influence the perception of overall facial colour appearance? In this section, three assumptions were brought up based on the three factors and their possible influence on the colour perception of facial appearance was discussed.

7.5.2.1 The outliners of the facial colours

The first assumption is that there might be an influence of the outliners of the facial colours. The skin colour is not uniform and the presence of skin

features such as wrinkles, acne, moles, pores, hairs, and the shadows and highlights generated by these things could add some unusual colour to the skin colour calculations. It is possible that when perceiving facial colour appearance, our human vision system would recognize these features and filter this colour information since they are not normal skin colours. If so, the trimmed mean colour of each pixel excluding the outlines would optimise the regression models and make a closer prediction of the perceived overall facial colour appearance. Based on that assumption, attempts were made to remove a designated percentage of the largest and smallest values before calculating the mean pixel colour of the faces, and then the colour difference between the matched colour and the trimmed mean colour was examined. The trimmed mean colour was calculated using the function *trimmean()* in MATLAB.

Figure 7.4 shows the average colour difference across eighty faces between the trimmed mean and the matched colour as a function of the percentage of the pixel colour used to calculate the mean after removing the outliners. It was found that the colour difference did decrease as more outliners are excluded from 3.14 ΔE^*_{ab} unit (between the mean of the middle 100% of the pixel colours and the matched colour) to 2.83 ΔE^*_{ab} unit (between the mean of the middle 60% of the pixel colours and the matched colour). The extreme value of the trimmed mean is actually the median of all pixel colours, which was 2.82 units of ΔE^*_{ab}. By removing the outliners and calculating the trimmed mean, the colour difference was reduced by 10.2%. The results indicated that the colour information of some outliners on skin appearance might be processed and filtered out by the vision system during the perception of facial colour appearance, but only to a relatively small extent. There might be other factors that influence the perception of the overall facial colour appearance.

$\triangle E_{a*b*}$ **between trimmed mean and the matched colour**

Percentage of the pixel colour after removing the outliers
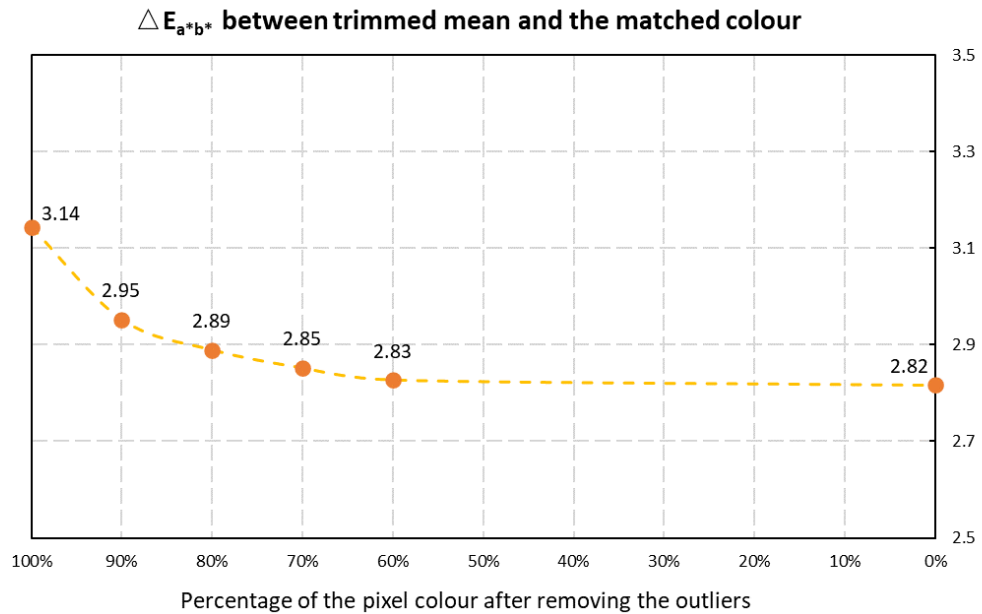
Figure 7.4 The ΔE_{a*b*} between the trimmed mean colour (excluding outliers) and the matched colour.

7.5.2.2 The skin colours of different local facial areas

Considering the particularity of skin colour, the way people observe a human face could be different from seeing any colour patches or other images. An earlier study showed when individuals examined facial displays, a greater amount of time was spent looking at the regions around the eyes as compared to the areas around the mouth (Janik et al., 1978). When assessing the overall colour appearance of facial skin, observers might also focus more on some important facial areas, instead of paying equal attention to the entire face. Thus, the second assumption is that the perceptual facial colour appearance is influenced by the local facial areas and is closer to the colour of the important areas of visual focus. In this respect, the colour of five local skin areas, forehead, cheek, nose, chin, and periorbital was considered, and their colour difference with both the average colour and the matched colour were examined.

Table 7.3 shows the average colour differences across all faces, Caucasian faces, and Chinese faces, respectively. On average, the matched colour was the closest to the colour of the forehead, followed by the cheek, chin, nose, and periorbital, while the average facial colour was the closest to the colour of the chin. The results were similar to both faces that the forehead would be the most suitable local area to represent the global colour percept or impression of a human face among the five facial areas.

Table 7.3 The $\Delta E_{a^*b^*}$ between the five local skin colours and the average pixel colour (the top three rows), the matched facial colour (the bottom three rows). The smallest colour differences in each row were marked as bold.

	Forehead	Cheek	Nose	Chin	Periorbital
The colour difference with the average colour					
Overall	3.22	4.82	3.55	**2.39**	5.11
CA Face	3.88	3.67	3.98	**2.25**	6.63
CN Face	2.56	5.97	3.12	**2.53**	3.60
The colour difference with the matched colour					
Overall	**2.09**	3.31	5.23	3.76	7.89
CA Face	**1.48**	2.79	6.25	3.78	9.41
CN Face	**2.70**	3.83	4.22	3.74	6.36

While the forehead colour showed the smallest colour difference with the matched facial colour on average, there was individual difference across different faces, especially for the forty Chinese faces (Figure 7.5). That was due to the individual difference in skin colour distribution of the face. The forehead might be a lighter region for most faces, but the colour of other local areas didn't show consistent a pattern across different faces, e.g. some faces had more reddish cheeks while others showed paler and lighter cheek colours, which resulted in the massive crosses of the lines in Figure 7.5. Therefore, in terms of the individual face, the selection of local areas representing the global facial colour impression is not determined and still depends on the colour distributions of that face.
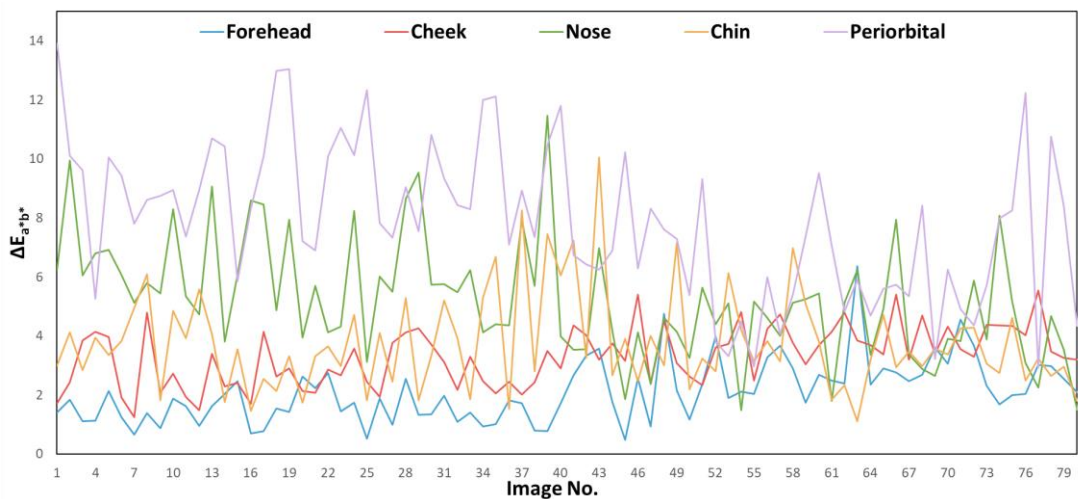
Figure 7.5 The $\Delta E_{a^*b^*}$ between the five local skin colours and the matched facial colour of each of the eighty faces (Image No.1-40: CA faces; No. 41-80: CN faces).

7.5.2.3 The colours of facial features

Although the observers were asked to make the colour match based on the overall skin colour, it was difficult to ignore the influence of the facial features (including eyes, brows, and mouth) on the skin colour perception. The perceptions of colour change according to the visual context, and such visual perception phenomenon refers to Simultaneous Contrast. According to the contrast effects on lightness, a test colour on a dark background appears lighter than the same colour on a light background (Luo et al., 1995). Comparing the average skin colour and the feature colours for both faces, it was found the colour of all three features, eyes, brows, and mouth, are much darker than the average skin colour (Table 7.4). Thus, the third assumption was made that the major colour shift in lightness (L*) was due to the contrast effects influenced by the feature colours. However, the current experiments are not enough to confirm whether this assumption is correct. A follow-up study may be required to test the difference in the matching results if only facial skin areas were presented excluding the facial features instead of the whole facial images.

Table 7.4 Descriptive statistics for the average facial skin colours and the facial feature colours in CA and CN faces.

	Mean (SD)		
	L*	a*	b*
CA Face			
Average	60.63 (2.06)	8.10 (1.26)	15.25 (2.06)
Eyes	39.80 (3.62)	5.76 (1.06)	10.60 (1.39)
Brows	42.94 (6.29)	7.94 (1.40)	13.51 (1.69)
Mouth	47.16 (2.29)	20.31 (2.31)	12.40 (1.19)
CN Face			
Average	55.92 (2.10)	8.72 (1.16)	18.55 (1.50)
Eyes	37.59 (2.93)	6.75 (0.98)	12.07 (1.11)
Brows	42.93 (5.24)	6.35 (1.30)	12.05 (1.99)
Mouth	44.02 (1.75)	19.02 (2.10)	13.59 (1.07)

7.5.3 The perceptual difference in facial colour appearance between Caucasian and Chinese observers

Interestingly, the difference in colour change between the observer of the two ethnic groups was significant in all three dimensions, $\triangle L*$ ($F_{1,42}=13.247$; $p<0.01$), $\triangle a*$ ($F_{1,42}=10.708$; $p<0.01$), and $\triangle b*$ ($F_{1,42}=13.479$; $p<0.01$). Both observers increased lightness (L*) to match the facial appearance, but Caucasian observers made more increase in lightness (L*), resulting in a

lighter appearance match compared to Chinese observers (Figure 7.2c). Both observers reduced redness (a*) and yellowness (b*) to match the overall facial colour appearance, but Chinese observers reduced more, resulting in a less yellowish and less reddish facial appearance match compared to Caucasian observers (Figure 7.2a and Figure 7.2b). Besides, the interaction effect between face ethnicity and observer ethnicity was found significant for the lightness change ($\triangle$L*, $F_{1,3239}$=18.143; p< 0.001) and the yellowness change (($\triangle$b*, $F_{1,3239}$=13.979; p< 0.001). Both observers tended to make more colour changes when they viewed faces of their own ethnicity. Caucasian observers made a larger increase in lightness (L*) when they viewed Caucasian faces, while Chinese observers tended to decrease more yellowness on the average colour to match the actual appearance colour when they viewed Chinese faces (Figure 7.2c and Figure 7.2b).

Such ethnic differences in facial colour perception have never been revealed before and possibly originated from the differences in the naturally occurring variations of skin chromatic properties or the long-term visual adaptation of the two ethnic groups. The systematic differences in the skin colour variations show that Caucasian faces have higher lightness (L*) and lower chroma (C*) values, while Chinese faces have lower lightness (L*) and higher chroma (C*) values (especially b*, see Figure 4.2). Such variations might contribute to the 'other-ethnicity effect' on skin colour perception that observers pay more attention to or have extra sensitivity in the colour dimensions (lightness or chroma) they are more familiar with. Tan et al. investigated the discrimination of facial colour change using Chinese participants and found the Chinese participants were more sensitive to the colour changes in redness (a*) and yellowness (b*), but not lightness (L*), yet observers of other ethnicity were not included in their study (Tan and Stephen, 2013). On the other hand, the perception of facial appearance can be strongly affected by the characteristics of faces viewed previously, and visual adaption can result in long-term changes, especially for more familiar faces (Webster et al., 2004; Webster and MacLeod, 2011). As Caucasian people are exposed to lighter faces on a daily basis while Chinese people are more familiar with their own faces with a bit higher chroma, the perception of the facial colour may be biased by prior exposure and linked to the observer's state of adaptation. Meanwhile, there is the 'other-ethnicity effect' for face adaptation that people are more sensitive to faces of their own ethnicity or faces that are close to the average faces to which there are adapted (Webster and MacLeod, 2011). In Tan et al.'s study, they also found that Chinese participants were significantly better at recognizing colour

differences in faces of their own and familiar than other-ethnicity faces (Tan and Stephen, 2013). This might partially explain the interaction effect that both observers tended to make more colour changes on faces of their own ethnicity.

Considering the above, further research is still needed to explore whether the perceptual differences between observers have come from their natural chromatic properties, state of adaption, sensitivity or other aspects.

7.6 Summary

In this chapter, the colour perception of overall facial skin appearance was investigated using a colour matching experiment. A summary of the analysis and major findings are given below:

- The average colour was not a good representation of the overall facial colour appearance.
- Overall, the actual facial colour appearance perceived by observers was less reddish, slightly less yellowish, and much lighter compared to the average facial colour which was commonly used in previous research.
- The overall facial colour appearance was accurately quantified and modelled for the first time by simple linear regressions of average pixel colour, and the models promised a high accuracy for facial colour appearance prediction with more than 95 per cent of the variance explained.
- Factors that influence the perception of overall facial colour appearance were discussed and three possible assumptions were brought up explaining why such global facial colour impressions formed.
- The perceptual difference in facial colour appearance between Caucasian and Chinese observers was further discussed.

Chapter 8 Facial whiteness, redness, and yellowness indices

8.1 Overview

In the previous chapter, the overall colour appearance of human faces was investigated by a colour matching experiment. A consistent tendency for the colour shift between the colorimetric average and the perceptual overall colour appearance was identified. While the overall colour appearance is important, skin whiteness, redness and yellowness are three attributes most directly describe people's perception of facial colour appearance and thus receive most concerns by industry applications such as cosmetic companies. The CIELAB is now widely used for skin colour, and the a* and b* values are commonly used to represent redness and yellowness, respectively. In this chapter, facial skin whiteness, redness, and yellowness were precisely examined. Whether the CIELAB coordinates (L*, a*, and b*) are suitable for quantifying those three perceptual attributes of facial skin colour, and whether there is a perceptual difference between facial skin colour and uniform patch colour were investigated.

Using the colour appearance scaling data collected in Experiment 4, statistical analysis (Section 8.2) was done to test first the relationship between the scaling results of the face and the patch with the matched colour appearance (Section 8.3), and then the relationship between perceived whiteness, redness, and yellowness and the CIELAB coordinates (Section 8.4). New indices of whiteness, redness and yellowness for facial skin were developed based on the multiple regression analysis (Section 8.5). The associations between skin whiteness, redness, and yellowness were also assessed (Section 8.6). Finally, the influence of L*, a*, and b* on skin colour appearance perception and the perceptual difference between face and patch were further discussed (Section 8.7).

8.2 Statistical analysis

The same 80 facial images were used in this chapter to study the perceived whiteness, redness, and yellowness of facial skin. The 80 corresponding uniform colour patches that have same appearance with each facial image (matched by a panel of observers in Experiment 3) were also used in this study to investigate whether facial feature would affect perception of facial whiteness, redness and yellowness appearance assessment. Each pair of the 80 faces and uniform colour patches were scaled regarding their perceptual whiteness, redness, and yellowness in Experiment 4. The scaling scores of facial skin whiteness, redness, and yellowness for each of the 80

facial images and from each of the 43 observers were recorded in the facial image session; the scaling scores of perceptual whiteness, redness, and yellowness for each of the 80 corresponding uniform colour patches and from each of the observer were recorded in the uniform patch session. The mean scaled scores over all observers for each face and each patch were then calculated to represent the perception of whiteness, redness, and yellowness by a panel of observers in the context of facial skin and uniform patches, respectively.

The scaling results of the face and the patch with the matched colour appearance were first compared using Pearson's correlation coefficients (two-tailed). The associations between the three perceptual attributes (whiteness, redness, and yellowness) and the CIELAB colorimetric values (L*, a*, and b*) were then assessed. Based on the associations, multiple linear regression models were set up to predict facial whiteness, redness, and yellowness from the L*, a*, and b* values, and the model performance was compared with the simple linear regressions. In addition, Pearson's correlation coefficients (two-tailed) were used to test the associations between the three perceptual attributes for the facial stimuli and the patch stimuli, respectively.

All the analyses were conducted in R (RDC, 2010). Pearson's correlation coefficients were calculated using the *cor()* function in the *stats* R package. The simple linear regression model and multiple regression models were built using the *lm()* function under the *stats* library.

8.3 Face colour vs. patch colour

The Pearson Correlation Coefficient (two-tailed) was used to assess the agreement in the scaling scores of the faces and the matched colour patches. The Pearson Correlation Coefficient of Caucasian (CA) faces and the corresponding patches, Chinese (CN) face and the corresponding patches, and all the faces and the corresponding patches are shown in the first, second, and third column in Table 8.1, respectively. The scaling scores of the faces and the patches were highly correlated for all three scaled attributes of whiteness, redness, and yellowness (p<0.001). The data points clustered around the 45° line in Figure 8.1 show the associations between the face scores and the patch scores and indicate a good agreement between them.

Table 8.1 The Pearson Correlation Coefficients of the scaling scores between the face stimuli and the matched patch stimuli.

	CA	CN	All
Whiteness	0.96***	0.96***	0.97***
Redness	0.91***	0.89***	0.88***
Yellowness	0.97***	0.93***	0.98***

*P<0.05; **P<0.01; ***P<0.001. N = 40, 40, 80 for CA, CN and all faces, respectively
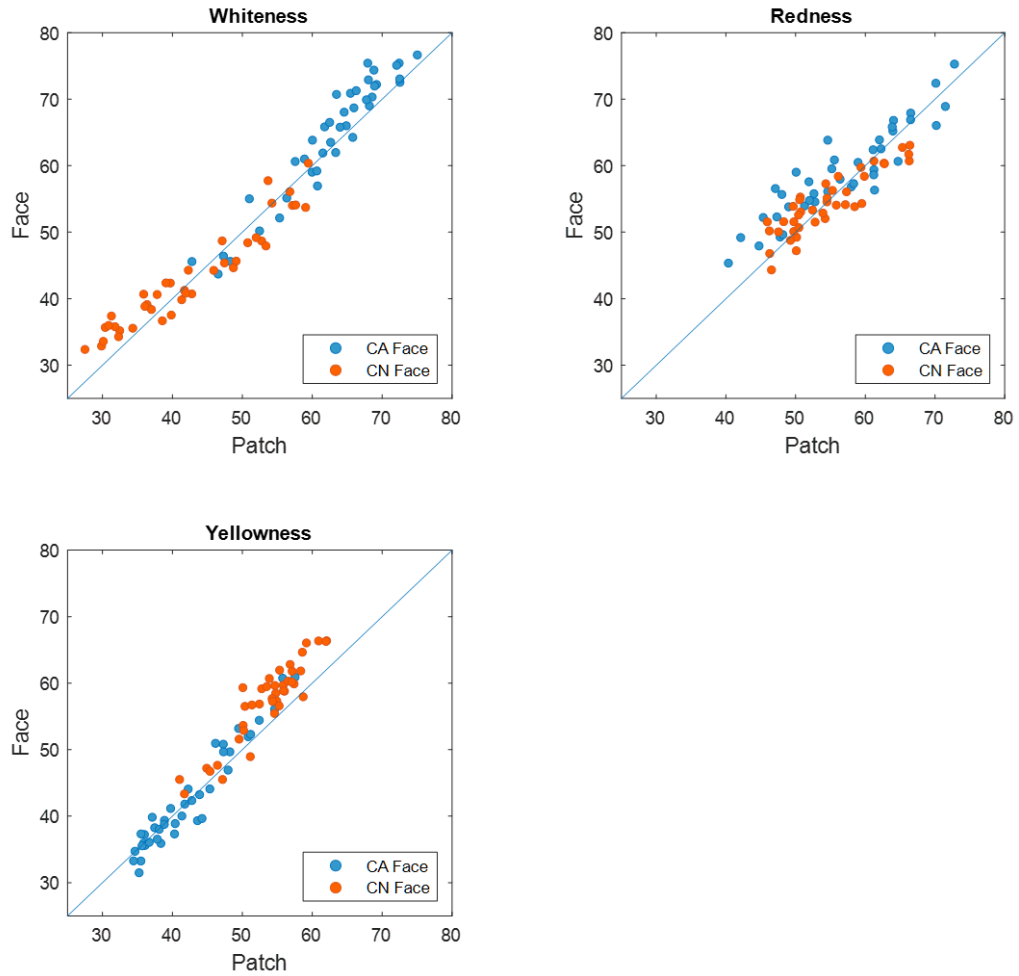
Figure 8.1 Associations between face scores and patch scores when scaled by three perceptual attributes, whiteness, redness, and yellowness: • Caucasian faces (CA), • Chinese faces (CN). A line has been drawn at 45° to facilitate comparison.

8.4 Whiteness, redness, and yellowness vs. L*, a*, and b*

The relationships between the three perceptual attributes (whiteness, redness, and yellowness) and the CIELAB colorimetric values (L*, a*, and b*) were investigated. Figure 8.2 and Figure 8.3 shows the correlations for the face stimuli and the uniform patch stimuli, respectively.

The perceived whiteness of facial skin showed stronger correlations with the L* and b* values than the a* value; Facial skin redness was found to have strong correlations with both the a* and b* values, and at a similar level; Facial skin yellowness was found most strongly correlated with the b* and also highly correlated with the L* value (Figure 8.2).
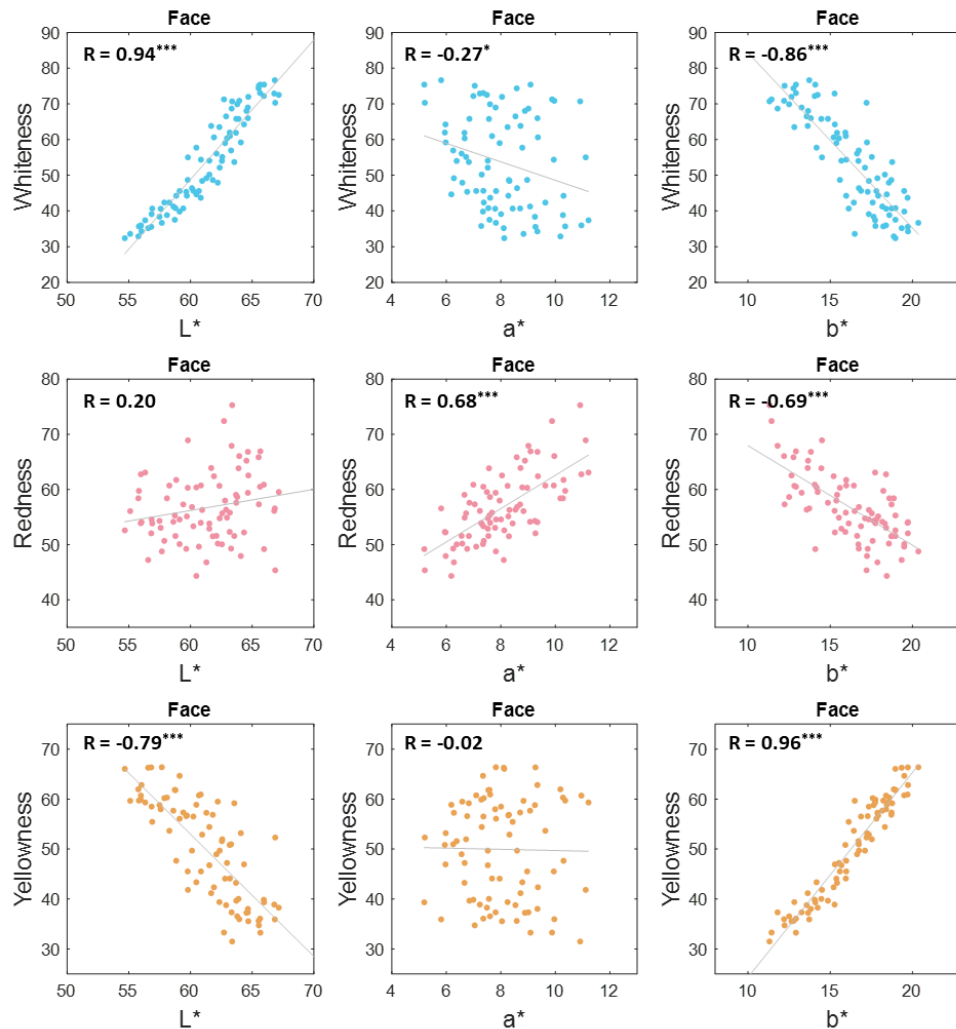
Figure 8.2 Correlations between the three perceptual attributes, whiteness, redness, and yellowness, and the three CIELAB coordinates, L*, a*, and b* of the 80 facial images. The Person correlation coefficients and the significance of the correlations are shown at the left top of each subplot. *P≤0.05, ** P≤0.01, ***P≤0.001.

The perceived whiteness of uniform skin patches also had a stronger correlation with the L* and b* values than the a* value; Patch redness was found more strongly correlated with the a* value than the b* value; Patch yellowness was found most strongly correlated with the b* and also correlated with the L* value (Figure 8.3).
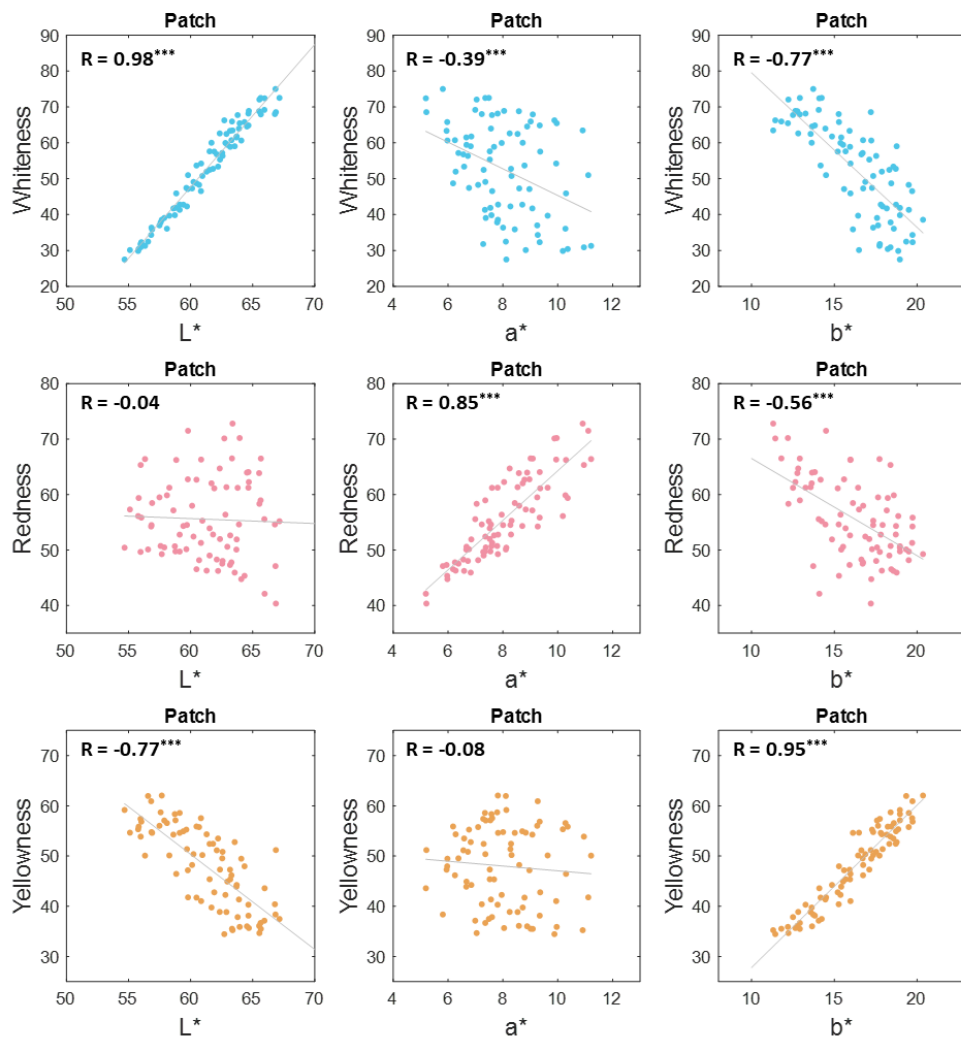
Figure 8.3 Correlations between the three perceptual attributes, whiteness, redness, and yellowness, and the three CIELAB coordinates, L*, a*, and b* of the 80 uniform colour patches. The Person correlation coefficients and the significance of the correlations are shown at the left top of each subplot. *P≤0.05, ** P≤0.01, ***P≤0.001.

8.5 Whiteness, redness, and yellowness indices for facial skin colour

Considering the perceived facial whiteness, redness, and yellowness were influenced not only by a single variable of the CIELAB coordinates, it was reasonable to assume that these perceptual colour attributes could be more accurately predicted by the L*, a*, and b* values together. Thus, multiple linear regression models were set up to predict facial whiteness, redness, and yellowness from the L*, a*, and b* values. For comparison, three simple linear regression models were also set up to predict facial whiteness from the L* value, facial redness from the a* value, and facial yellowness from the

b* value. Table 8.2, 8.3 and 8.4 summarises the regression analysis for the perceived facial whiteness, redness, and yellowness, respectively. Compared to the simple linear regression, the multiple regression models explained more variance in the prediction of all three perceptual colour attributes (98.2, 87.4, and 96.8 per cent of the variance in facial whiteness, redness, and yellowness, respectively, was explained by the multiple regression models). Especially for the prediction of perceived facial redness, the model fit was largely improved from an R^2 value of 46.1% (simple regression) to an R^2 value of 87.4% (multiple regression).

Table 8.2 Simple linear regression model and multiple linear regression model in predicting facial whiteness.

	Facial Whiteness			
	β	SE	*t*	P
Regression 1: Whiteness ~ L*				
Model	$F_{1,78}$=620; P<0.001***; Adjusted R^2=0.887			
(Intercept)	-185.784	9.631	-19.29	<0.001***
L*	3.910	0.157	24.9	<0.001***
Regression 2: Whiteness ~ L*, a*, b*				
Model	$F_{3,76}$=1405; P<0.001***; Adjusted R^2=0.982			
(Intercept)	-75.117	10.455	-7.185	<0.001***
L*	2.743	0.118	23.211	<0.001***
a*	-0.058	0.205	-0.285	0.776
b*	-2.385	0.146	-16.348	<0.001***

NS= not significant; *P<0.05; **P<0.01; ***P<0.001.

Table 8.3 Simple linear regression model and multiple linear regression model in predicting facial redness.

	Facial Redness			
	β	SE	*t*	P
Regression 1: Redness ~ a*				
Model	$F_{1,78}$=68.67; P<0.001***; Adjusted R^2=0.461			
(Intercept)	32.375	2.975	10.884	<0.001***
a*	3.020	0.364	8.287	<0.001***
Regression 2: Redness ~ L*a*b*				
Model	$F_{3,76}$=183.6; P<0.001***; Adjusted R^2=0.874			
(Intercept)	24.415	12.535	1.948	0.055
L*	0.424	0.142	2.995	0.004**
a*	3.281	0.246	13.361	<0.001***
b*	-1.239	0.175	-7.087	<0.001***

NS= not significant; *P<0.05; **P<0.01; ***P<0.001.

Table 8.4 Simple linear regression model and multiple linear regression model in predicting facial yellowness.

	Facial Yellowness			
	β	SE	*t*	P
Regression 1: Yellowness ~ b*				
Model	$F_{1,78}$=897.4; P<0.001***; Adjusted R²=0.919			
(Intercept)	-16.671	2.246	-7.421	<0.001***
b*	4.100	0.137	29.957	<0.001***
Regression 2: Yellowness ~ L*a*b*				
Model	$F_{3,76}$=794.2; P<0.001***; Adjusted R²=0.968			
(Intercept)	84.563	10.316	8.198	<0.001***
L*	-1.216	0.117	-10.429	<0.001***
a*	-0.960	0.202	-4.749	<0.001***
b*	2.928	0.144	20.346	<0.001***

NS= not significant; *P<0.05; **P<0.01; ***P<0.001.

Figure 8.4 shows the model performance of the simple linear regression model and the multiple regression model. Generally, the multiple regression model on the right column shows a lower level of dispersion of the data points from the diagonal line and suggests a better performance of the model compared to the simple regression model on the left column.

Hence, a new Whiteness Index for Skin (WIS), a new Redness Index for Skin (RIS), and a new Yellowness Index for Skin (YIS) were developed based on the multiple regression analysis to accurately quantify and predict the perceived facial whiteness, redness, and yellowness, respectively, from the CIELAB coordinates. The formulas are given below:

$$\text{WIS (Whiteness Index for Skin)} = 2.743\bar{L}^* - 0.058\bar{a}^* - 2.385\bar{b}^* - 75.117;$$

$$\text{RIS (Redness Index for Skin)} = 0.424\bar{L}^* + 3.281\bar{a}^* - 1.239\bar{b}^* + 24.415;$$

$$\text{YIS (Yellowness Index for Skin)} = -1.216\bar{L}^* - 0.960\bar{a}^* + 2.928\bar{b}^* + 84.563.$$

Equation 8.1
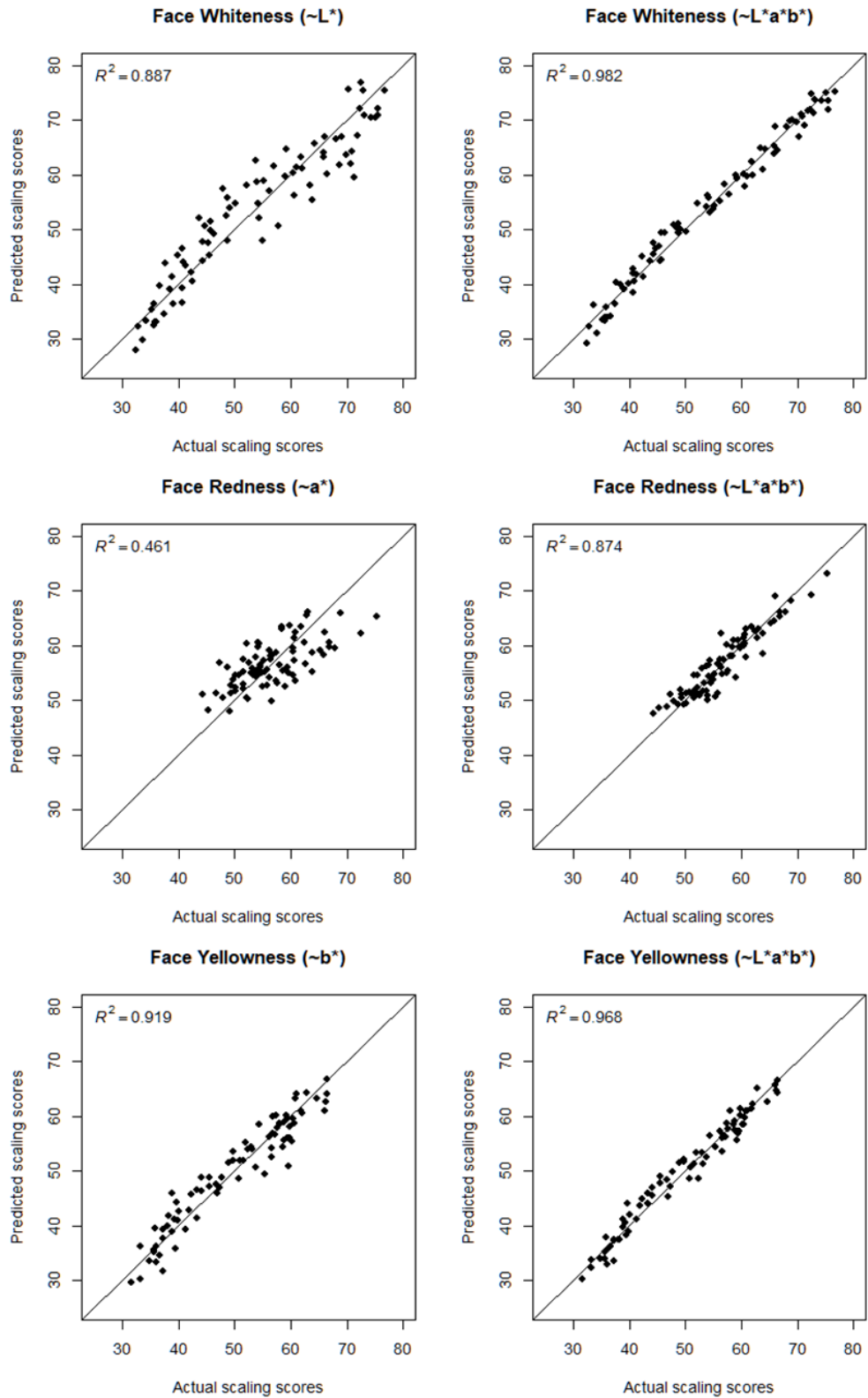
Figure 8.4 The model performance of the simple regression models (left column) and multiple regression models (right column) in predicting facial whiteness (top row), redness (middle row), and yellowness (bottom row).

8.6 Associations between perceptual whiteness, redness, and yellowness

At last, the Pearson Correlation Coefficient (two-tailed) was used to assess the associations between the three perceptual colour attributes, whiteness, redness, and yellowness. Table 8.5 shows the results of the face stimuli and the uniform patch stimuli. Within the constraints of the skin colour space, no matter whether the face or the uniform patch was scaled, the perceived whiteness, redness, and yellowness were not independent colour perceptions. Especially, whiteness and yellowness were strongly and negatively correlated perceptions (r = -0.93, -0.84 for face stimuli and patch stimuli, respectively). Comparing the two different stimuli, the face and the patch, higher correlations were found between any two of the three perceptual attributes in the face stimuli than the patch stimuli. The results indicated the colour perception of human faces could be more easily affected by perceptions of other colour dimensions compared to the uniform colour patches. In particular, perceived whiteness and redness were significantly correlated when evaluating the facial skin colour (p<0.001), whereas they were not related to each other when scaling the uniform skin colour patches (p>0.05).

Table 8.5 The Pearson Correlation Coefficients of whiteness, redness, and yellowness for face and patch perceptions.

	Face	Patch
Whiteness - Redness	0.42***	0.06
Whiteness - Yellowness	-0.93***	-0.84***
Redness -Yellowness	-0.65***	-0.54***

*P<0.05; **P<0.01; ***P<0.001. N=80, 80 for face and patch, respectively

8.7 Discussion

8.7.1 New indices of whiteness, redness, and yellowness for facial skin colour

This study aimed to investigate (1) the relationship between perceived skin whiteness, redness, and yellowness and the colorimetric values, L*, a*, and b*, of the facial skin. (2) the perceptual differences between face stimuli and the uniform patch stimuli with the matched colour appearance. A psychophysical experiment was conducted using the method of magnitude estimation, and both the 80 real facial images and the corresponding 80 uniform skin colour patches were scaled by three perceptual attributes,

whiteness, redness, and yellowness. Based on the multiple regression analysis, the three new indices, WIS, RIS, and YIS, were developed to accurately quantify and predict the perceived whiteness, redness, and yellowness of human faces.

For a long time, the L*, a *, and b* values in CIELAB space have been taken for granted as the representatives of facial lightness, redness, and yellowness, respectively, in a large number of studies on facial colour preferences and facial impressions (Stephen, Law Smith, et al., 2009; Re et al., 2011; Stephen et al., 2011; Pazda et al., 2016; Jones et al., 2016; Foo, Simmons, et al., 2017; Thorstenson et al., 2017; Han et al., 2018; Jones, 2018). Perceived facial whiteness was found highly associated with lightness (L*) and is particularly important in Asian beauty (Xie and Zhang, 2013; Gao et al., 2018; Shimakura and Sakata, 2019). The results of this study suggested that the perceptual facial whiteness was not simply a correlate of CIELAB L*, the perceptual facial redness was not simply a correlate of CIELAB a*, and the perceptual facial yellowness was not simply a correlate of CIELAB b*. Especially for the redness, it was proved that both the a* value and the b* value correlated with facial redness (r = 0.68 and 0.69, both p<0.001, see Figure 8.2), and the a* value was not a good predictor of skin redness (R^2 = 0.461, see Table 8.3). In this study, it was the first time that the accurate relationships between perceived skin whiteness, redness, and yellowness and the colorimetric values, L*, a*, and b* were revealed as given by the new indices WIS, RIS, and YIS. The stronger performance of WIS, RIS, and YIS relative to the simple regressions was expected given that facial whiteness had a strong correlation with both L* and b*, facial redness had a strong correlation with both a* and b*, and facial yellowness had a strong correlation with both L* and b*.

Previous whiteness metrics, the ITA° scale and the depth scale D^*_{ab} were tested and both showed a good correlation with the facial whiteness scaling scores from the experiment (Figure 8.5), and both were better than the simple regression of L* (R^2 = 0.887, see Table 8.2). Whereas the new WIS developed based on the visual assessment data of facial skin colour outperformed both the ITA° scale and the depth scale D^*_{ab} in predicting facial whiteness. Facial redness and facial yellowness have never been quantified previously in the context of skin colour space. The new RIS, YIS, together with WIS developed in this study can be used as accurate measures of perceptual facial colour appearance in various applications as well as the studies related to skin colour appearance.
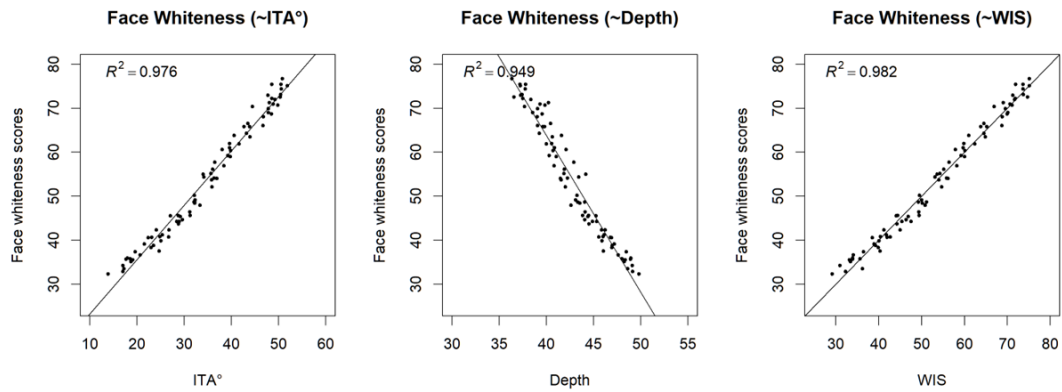
Figure 8.5 Perceived facial whiteness as a function of the ITA scale (left), Bern's Depth scale (middle), and the new WIS scale (right).

8.7.2 Effects of a* and b* (h_{ab} and C^*_{ab}) on whiteness perceptions

The results suggested the perceptions of whiteness, redness, and yellowness of both the face stimuli and the patch stimuli were affected by more than one single dimension of L*, a*, and b*. As the images of real human faces were used in this study, it could be difficult to examine the effect of each variable precisely and independently. In fact, the effects of L*, a*, and b* on the perceptual colour attributes could be affected by the intrinsic correlations between L*, a*, and b* as they were not independent variables within the real skin colour gamut (e.g. L* correlated with both a* and b*, see Table 5.2). In the following discussions, efforts were made to divide the L*, a* or b* values into subgroups, and to show, at least partially, the influence trend of L*, a*, and b* on the perceptual colour attributes as well as the perceptual differences between the face stimuli and the patch stimuli within the same subgroups.

To discuss the influence of a* and b* on whiteness perceptions, the L* value was divided into five groups by a step of 2.5 L* units so that within each group the L* value was kept at an approximately similar level. Figure 8.6 shows the results of the face stimuli and the patch stimuli side by side, and a regression line was drawn for each group. Table 8.6 at the end of this section summarises the slopes of all the regression lines for the face stimuli and the patch stimuli, respectively. The top row of Figure 8.6 shows the perceived whiteness of the face stimuli was more sensitive to the change of a* compared to that of the patch stimuli (slopes of the face were steeper than the patch in all groups, see Table 8.6). And a higher a* value was more easily to increase the perceived whiteness of faces in the higher L* range (slopes of the higher L* group were steeper than lower L* group). The bottom row of Figure 8.6 shows that a higher b* value decreased the

perceived whiteness of both the face stimuli and the patch stimuli (negative slopes), whereas facial skin whiteness was more sensitively affected by b* (slopes of the face were steeper than the patch in all groups, see Table 8.6). Meanwhile, a higher b* value was more easily to decrease the perceived whiteness of faces in the higher L* range (slopes of the higher L* group were steeper than the lower L* group).

Figure 8.6 Relationships between whiteness scores and a* (top), and whiteness scores and b* (bottom) for facial colour perceptions (left) and patch colour perceptions (right). A regression line was drawn for each group.

Similar results were found by Yoshikawa et al. who studied the effects of chromatic components on facial skin whiteness. Yoshikawa et al. revealed that the perceived facial skin whiteness was influenced by both the hue and chroma that reddish facial skin looked whiter than yellowish skin and low-chroma skin looked whiter than high-chroma skin (Yoshikawa et al., 2012). Shimakura and Sakata also revealed facial skin image looked less bright with an increased chroma (the inverse Helmholtz–Kohlrausch effect) whereas a uniform colour patch looked brighter with an increase in chroma (the Helmholtz–Kohlrausch effect) (Shimakura and Sakata, 2019). In such cases, the effects of hue and chroma on whiteness perception in this study

were also investigated as Figure 8.7 shows. The top two plots show that reddish skin looked whiter than yellowish skin (negative slopes) and perceived facial whiteness was more sensitive to the hue change than the patch stimuli (slopes of the face were steeper than the patch in all groups, see Table 8.6). The bottom two plots show the inverse Helmholtz–Kohlrausch effect for both the face stimuli and the patch stimuli (negative slopes in all groups except for group 1) and the face stimuli again showed greater sensitivity to the chroma change (slopes of the face were steeper than the patch in all groups except for group 1, see Table 8.6). Different from Shimakura and Sakata's study, the Helmholtz–Kohlrausch effect was not found in the perception of uniform colour patches. This might be due to the face shape of the uniform patch stimuli used in this study, which could to some extent evoke the process of face recognition but the degree was weaker than the real face perception.
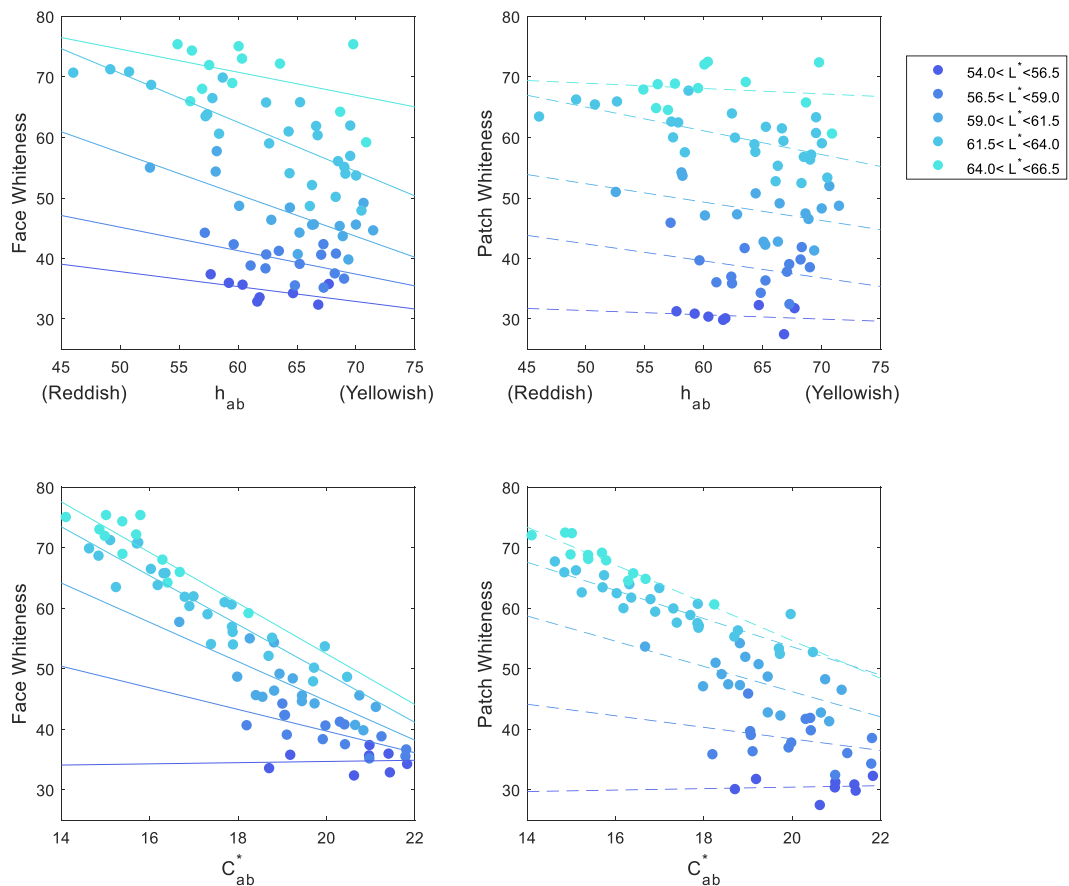
Figure 8.7 Relationships between whiteness scores and h_{ab} (top), and whiteness scores and C^*_{ab} (bottom) for facial colour perceptions (left) and patch colour perceptions (right). A regression line was drawn for each group.

Table 8.6 Summary of the slopes of the regression lines in Figure 8.6 and Figure 8.7.

Group	L* Range	Whiteness - a*		Whiteness -b*	
		Face	Patch	Face	Patch
1	54-56.5	0.57	0.19	-0.48	-0.03
2	56.5-59	0.59	0.53	-1.56	-0.93
3	59-61.5	2.15	0.70	-2.53	-1.43
4	61.5-64	3.24	1.28	-2.91	-1.58
5	64-66.5	0.64	-0.40	-3.11	-1.86

Group	L* Range	Whiteness - h_{ab}		Whiteness -C*	
		Face	Patch	Face	Patch
1	54-56.5	-0.25	-0.07	0.10	0.12
2	56.5-59	-0.39	-0.28	-1.79	-0.96
3	59-61.5	-0.69	-0.30	-3.25	-2.09
4	61.5-64	-0.81	-0.39	-4.04	-2.33
5	64-66.5	-0.38	-0.09	-4.19	-3.13

8.7.3 Effects of L*, b* on redness perceptions and L*, a* on yellowness perceptions

The a* value was also divided into five groups by a step of 1 a* unit to demonstrate the relationships between redness perceptions and the L* and b* values (Figure 8.8). The top row in Figure 8.8 shows that within each group both facial skin redness and patch redness were influenced by L*, and a higher L* value increased the perceived redness (positive slopes except for group 1). The perceived redness of the face stimuli was more sensitive to the change of L* compared to that of the patch stimuli (slopes of the face were steeper than the patch in all groups, see Table 8.7). The effects of b* on redness perception were similar between the face stimuli and the patch stimuli that a higher b* value decreased the perceptual redness of both the facial skin and the uniform patch (negative slopes).
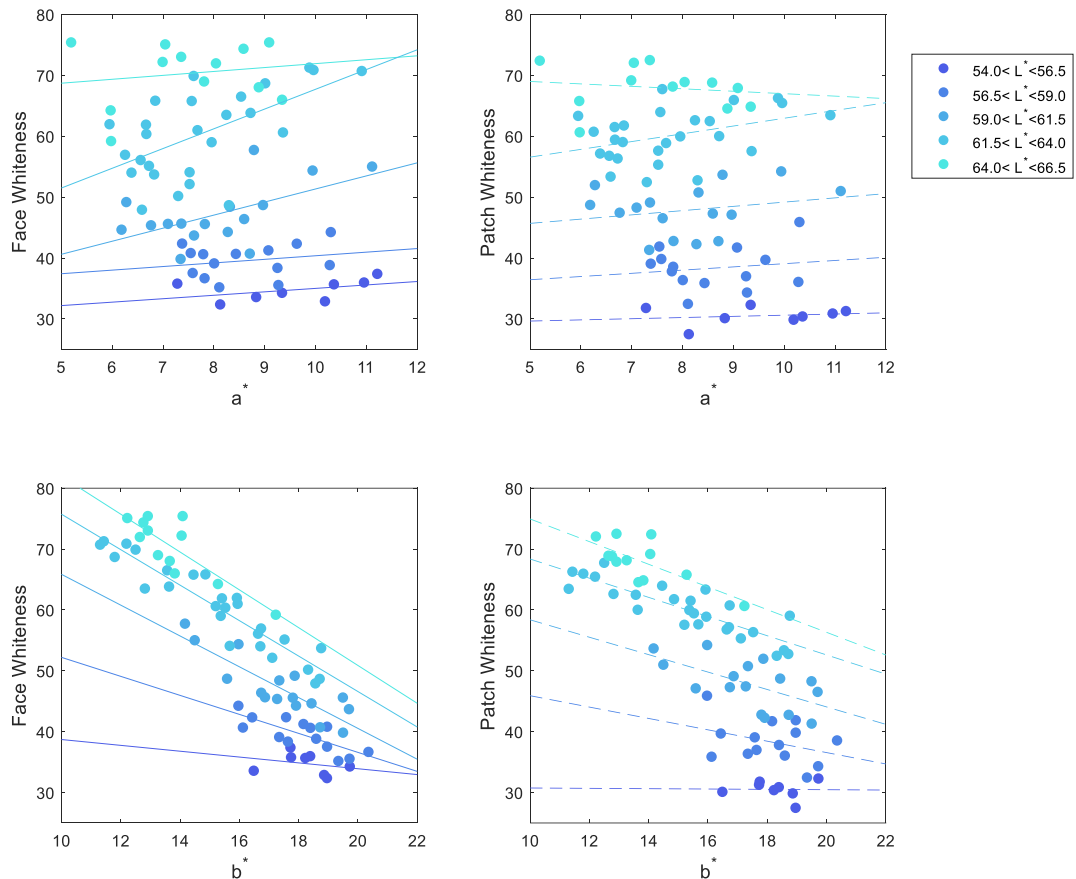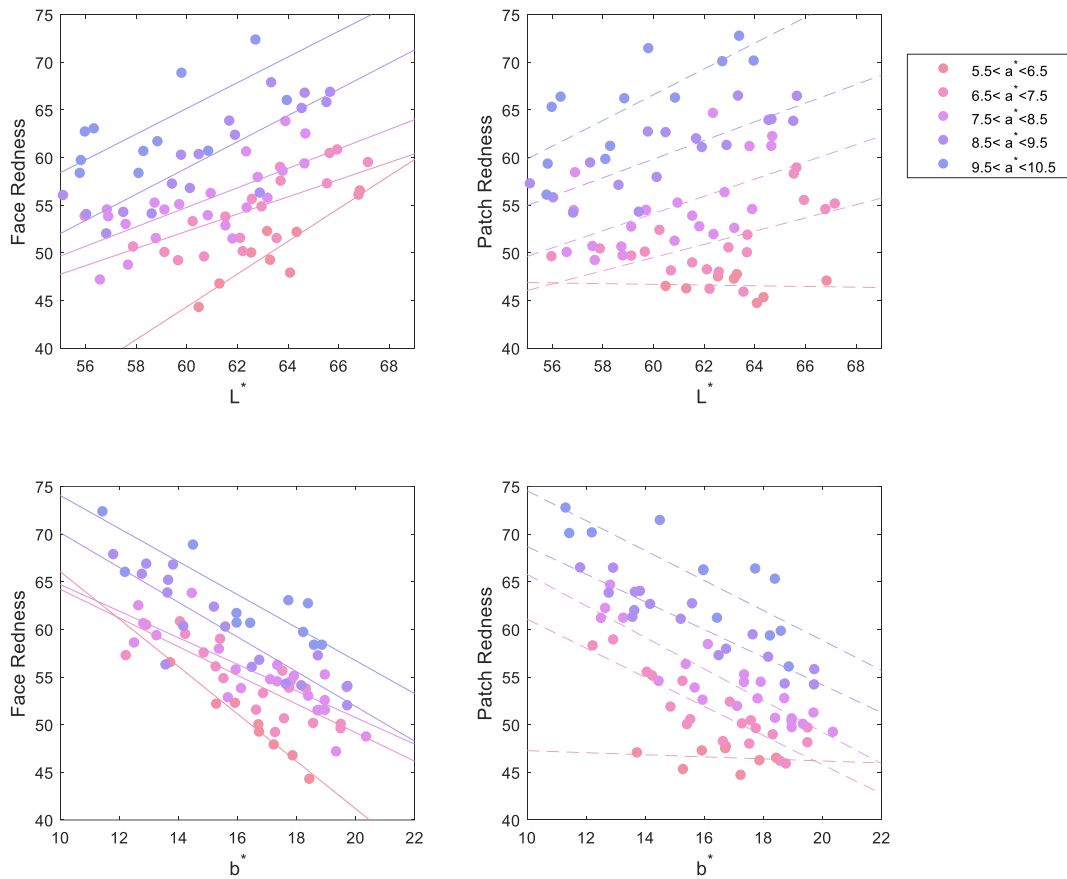
Figure 8.8 Relationships between redness scores and L* (top), and redness scores and b* (bottom) for facial colour perceptions (left) and patch colour perceptions (right). A regression line was drawn for each group.

At last, the b* value was divided into five groups by a step of 2 b* units to illustrate the effects of L* and a* on yellowness perceptions (Figure 8.9). The top plots in Figure 8.9 show a higher L* value decreased the perceptual yellowness of both the face stimuli and the patch stimuli (negative slopes), whereas the facial skin yellowness was more sensitively influenced by the L* (slopes of the face were steeper than the patch in all groups, see Table 8.7). The influence of the a* on yellowness perceptions didn't show a consistent trend across different groups.
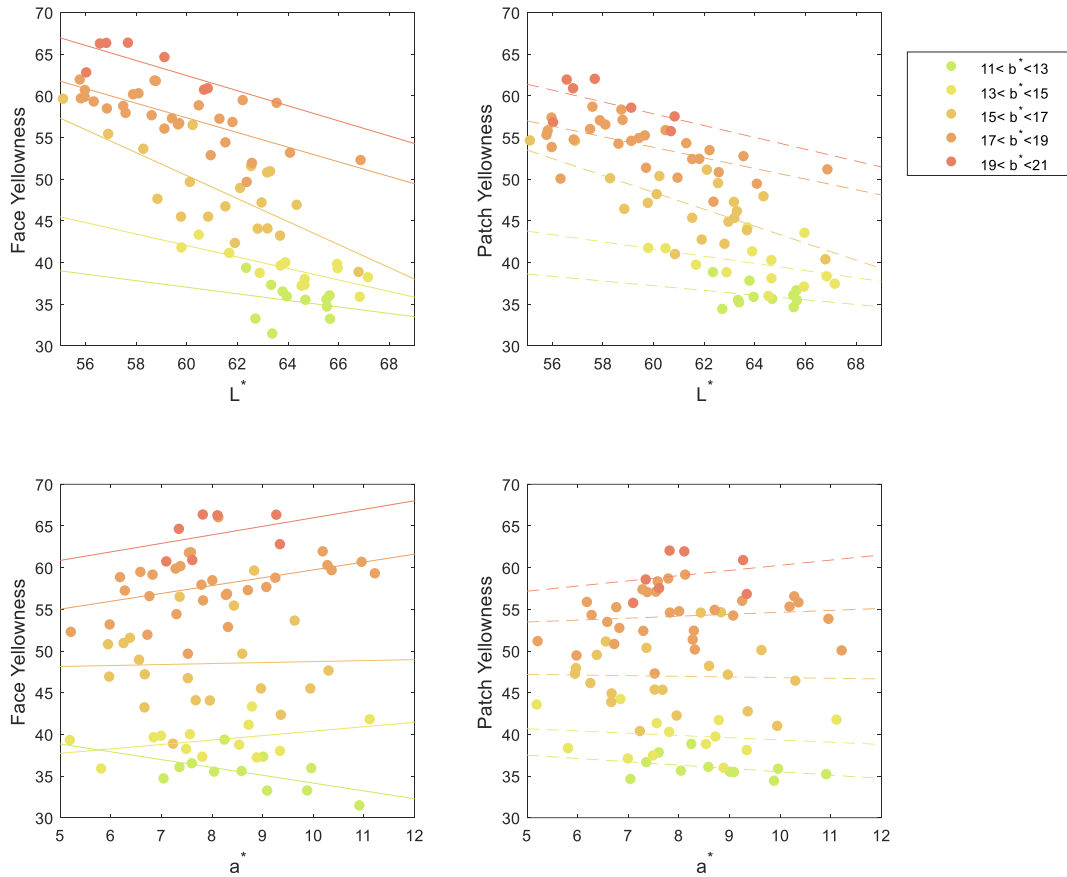
Figure 8.9 Relationships between yellowness scores and L* (top), and yellowness scores and a* (bottom) for facial colour perceptions (left) and patch colour perceptions (right). A regression line was drawn for each group.

Table 8.7 Summary of the slopes of the regression lines in Figure 8.8 – Figure 8.9.

Group	a* Range	Redness - L*		Redness - b*	
		Face	Patch	Face	Patch
1	5.5-6.5	1.72	-0.04	-2.49	-0.11
2	6.5-7.5	0.90	0.69	-1.50	-1.53
3	7.5-8.5	1.02	0.90	-1.40	-1.66
4	8.5-9.5	1.38	0.98	-1.82	-1.45
5	9.5-10.5	1.35	1.35	-1.73	-1.57

Group	b* Range	Yellowness - L*		Yellowness - a*	
		Face	Patch	Face	Patch
1	11-13	-0.39	-0.28	-0.94	-0.40
2	13-15	-0.69	-0.43	0.53	-0.27
3	15-17	-1.38	-1.02	0.12	-0.08
4	17-19	-0.88	-0.64	0.94	0.23
5	19-21	-0.91	-0.71	1.02	0.62

8.7.4 Perceptual difference between face and patch

While the scaling results between the face stimuli and the patch stimuli with the matched appearance were highly correlated (Section 8.3), perceptual differences were found between these two stimuli. The associations between the three perceptual colour attributes, whiteness, redness, and yellowness showed that higher correlations existed between any two of the three attributes in facial skin colour perceptions than the patch colour perceptions (Section 8.6). This might suggest that perceived whiteness, redness, and yellowness of human faces are more dependent on each other whereas when evaluating the uniform skin colour patches, these perceptual colour attributes could be easier to scale separately. The discussions in Section 8.7.2 and Section 8.7.3 also indicated that perceptual differences existed between the face stimuli and the patch stimuli. Steeper slopes were normally found in face stimuli indicating the greater sensitivity of facial colour perception which might be easier affected by the change of L^*, a^*, or b^* compared to the colour perceptions of uniform colour patches. These results were consistent with the associations between perceptual whiteness, redness, and yellowness in Section 8.6. Consistent with the previous research (Yoshikawa et al., 2012; Shimakura and Sakata, 2019), the perceptual differences between the face stimuli and the patch stimuli might indicate the higher-level process of face recognition which has an impact on colour perceptions.

In general, whiteness, redness, and yellowness were not independent perceptual attributes within the context of facial skin colour. The perceived facial whiteness could be enhanced by increased facial redness or a^*, but weakened by increased facial yellowness or b^*, and vice versa. The perceived facial redness could be weakened by increased facial yellowness or b^* and vice versa.

8.8 Summary

In this chapter, the perceptual whiteness, redness, and yellowness of both the face stimuli and the patch stimuli were investigated using the method of magnitude estimation. A summary of the analysis and major findings are given below:

- Psychophysical data was collected to quantitatively model the perceptual attributes of skin colour appearance.

- The accurate relationships between perceived skin whiteness, redness, and yellowness and the CIELAB colorimetric values, L*, a*, and b* were revealed.

- New indices WIS, RIS, and YIS were developed to quantify facial skin's perceptual whiteness, redness, and yellowness, respectively. Based on the multiple regressions, the new indices promised a high accuracy of prediction (R^2 = 0.982, 0.874, and 0.968 for WIS, RIS, and YIS, respectively).

- The perceptual differences between the face stimuli and the patch stimuli were discussed.

- In general, the perceived whiteness, redness, and yellowness were not independent perceptual attributes within the context of facial skin colour, and were more likely to influence each other compared to the perceptions of uniform patches.

Chapter 9 Conclusions

9.1 Overview

This research aims to better understand the human colour perception of facial complexions and to develop colour science models for colour preference evaluation and colour appearance perception. To achieve this aim, four psychophysical experiments were carried out in this research, assessing facial colour preference and facial colour appearance based on realistic skin models.

Experiment 1 was a cross-cultural study on facial preference evaluation. Eighty facial images of real human faces, including 40 Caucasian faces and 40 Chinese faces were used as stimuli in this experiment. Observers from the two ethnic groups were asked to rate the skin colour of each face, in terms of three preference attributes, facial attractiveness, perceived healthiness and visual ageing using the categorical judgement method.

Experiment 2 was a repeating experiment on facial preference evaluation using a new set of Chinese faces and a new panel of Chinese observers. The experimental procedures were the same as Experiment 1. The skin colour of the same 40 Chinese images used in Experiment 1 and another 60 new Chinese images was rated in terms of facial attractiveness using the categorical judgement method.

Experiment 3 was a colour appearance matching experiment focusing on the overall colour perception of facial appearance. The eighty facial images of real human faces were again used. Observers were asked to adjust the colour of the face-shaped uniform colour patch to match the skin colour appearance of each of the faces.

Experiment 4 was a colour appearance scaling experiment considering the perceived facial whiteness, redness, and yellowness, and the perceptual difference between the face stimuli and the uniform patch stimuli. The stimuli included eighty facial images and eighty uniform skin colour patches. The perceived whiteness, redness, and yellowness of both the face stimuli and the uniform patch stimuli were scaled in comparison to a reference using the magnitude estimation technique.

Based on these experiments, the research work was further done by analysing the role of average skin colour in preference judgements (Chapter 4), analysing the role of various facial colour characteristics in preference judgements (Chapter 5), providing an analytical framework and suggested regression techniques for facial attractiveness modelling (Chapter 6),

developing models to quantify and predict the overall colour appearance of human faces (Chapter 7), and developing new indices to quantify and predict facial skin's perceptual whiteness, redness, and yellowness (Chapter 8).

All the tasks described in Chapter 1 have been successfully achieved. A more detailed summary of the major findings and research contributions is given in the following sections. The directions for future study is also discussed in the end.

9.2 The role of facial colour cues in preference judgements based on realistic skin models

Based on the results of Experiment 1, the effect of various facial colour cues on facial preference judgements was thoroughly assessed and the relative importance of different colour cues was revealed. In order to discuss the facial colour preference based on a realistic skin model, a set of high-resolution images of real faces was used without changing the original colour and a rigorous process of display colour characterisation was performed to truly present the colour appearance of those facial images to observers in the preference evaluation experiments. The results underline the role of facial colour in nonverbal social communication such as determining facial preference. Although the average skin colour of facial areas plays a limited role, together with colour variation and contrast, there are stronger links between colour and facial preference than previously revealed. The results highlight the importance of examining various facial colour cues to obtain the full picture of colour predictors utilised in facial preference evaluation. The present study also contributes to the growing body of work demonstrating the importance of skin colour manipulations within an evolutionary meaningful parameter space, ideally using realistic skin models based on physical parameters.

9.3 The cultural difference regarding facial colour preference between Caucasian and Chinese populations

Meanwhile, interesting cultural differences regarding facial colour reference between Caucasian and Chinese populations were found. Facial colour characteristics as cues for attractiveness, healthiness, and youthfulness are deployed differently by Caucasian and Chinese observers. For example, the preference for facial lightness is opposite that Caucasian people prefer

tanning skin while Chinese people prefer lighter or whiter skin; Caucasian observes heavily rely on skin colour variation to make preference judgements, whereas facial colour contrasts are more important factors to Chinese observers; Chinese observers tend to rely more heavily on colour cues to judge all facial preference attributes than Caucasian observers; Caucasian observers think that both a younger face and older face can have an attractive and healthy appearance, whereas Chinese observers think younger faces mean healthier look and are more attractive; etc. From a psychological perspective, this is a very important subject that involves cross-cultural differences in human behaviour. Also, it is a subject that commands close attention by the industry such as cosmetics from both a colour and beauty perspective.

9.4 A new analytical framework for facial preference modelling

Attractiveness modelling could be complicated due to a large number of explanatory variables and correlations between them. In order to verify the out-of-sample performance of different modelling techniques, Experiment 2 was conducted to collect an independent testing dataset using new face stimuli and new observers. Results show that both dimension reduction regression and regularisation regression outperform the classical multiple and stepwise regressions, and the elastic net regression shows the most accurate predictions with relatively fewer colour variables selected. More importantly, the present study provides a useful and repeatable analytical framework for facial preference modelling based on a realistic skin model. The framework includes steps of variable identification, pre-processing, modelling, and model performance comparisons. It can serve as an important analytical tool in future studies of facial preference modelling and prediction from a large number of explanatory variables.

9.5 The overall colour appearance of human faces

It is the first time that the overall colour perception or the global colour impression of facial appearance is precisely examined. The results of Experiment 3 show that the perceived facial colour appearance is less reddish, slightly less yellowish, and much lighter compared to the average facial colour which was commonly used in previous research to represent the overall facial colour appearance. The overall facial colour appearance is accurately quantified and modelled by simple linear regressions of the

average pixel colour, and the models promise a high accuracy for facial colour appearance prediction with more than 95 per cent of the variance explained. Factors that may influence the perception of overall facial colour appearance were discussed. The perceptual difference in facial colour appearance between Caucasian and Chinese observers was interestingly found.

9.6 New indices of facial skin whiteness, redness, and yellowness

Results of Experiment 4 show that the perceived facial whiteness, redness, and yellowness cannot be simply represented by the L*, a*, and b* values in CIELAB colour space, respectively. Especially, the a* value is not a good predictor of skin redness (R^2 = 0.461). Based on the relationships between perceived skin whiteness, redness, and yellowness and the CIELAB colorimetric values, new indices WIS, RIS, and YIS are developed to accurately quantify those colour perceptions, which promise a high accuracy of prediction (R^2 = 0.982, 0.874, and 0.968 for predicting facial whiteness, redness, and yellowness from WIS, RIS, and YIS, respectively). The new indices can be used as accurate measures of perceptual facial colour appearance in various applications as well as studies related to skin colour appearance. In general, whiteness, redness, and yellowness are not independent perceptual attributes within the context of facial skin colour. The perceived facial whiteness can be enhanced by increased facial redness or a*, but weakened by increased facial yellowness or b*, and vice versa. The perceived facial redness can be weakened by increased facial yellowness or b* and vice versa.

9.7 The perceptual difference between the colour appearance of the face stimuli and the nonface stimuli

Meanwhile, the results of Experiment 4 also confirmed the perceptual differences between the face stimuli and the patch stimuli. Within the range of skin colour, the perceived whiteness, redness, and yellowness are not independent perceptual attributes no matter for the face stimuli or the uniform patch stimuli. However, higher correlations exist between any two of the three attributes in facial skin colour perceptions than the patch colour perceptions, showing these three colour appearance attributes are more likely to influence each other for the face stimuli compared to the uniform

patch stimuli. The perceptual differences may involve the influence of a higher-level process of face recognition on colour perceptions.

9.8 The direction for future study

The present study shed some new light on how our visual system perceives and processes colour information on human faces. Skin colour and face perception is a relatively new and multi-disciplinary research area covering colour technology, vision science, psychology, human behaviour, etc. and shaping numerous industry applications. There are various promising directions for future work of this research area.

The effect of various facial colour cues on preference judgements has been thoroughly examined in this study, and all the facial colour characteristics were calculated in the CIELAB uniform colour space. Considering that the L^*, a^*, and b^* are not perfect predictors of the perceptual facial whiteness, redness, and yellowness, respectively, the CIELAB may not be the best colour appearance model to describe the human colour perception of complexions and to predict facial preference judgements. Though the CIELAB uniform colour space is the most widely adopted colour space to objectively specify skin colours in numerous research, it was developed based on the uniform patch colours rather than skin colours as all the existing colour appearance models were. Given the perceptual difference between the face stimuli and the uniform patch stimuli, the current colour appearance model may not be adequate for the prediction of the much more complicated appearance of human skin or face. Based on the perception of skin colour, a better colour appearance model for human complexions could be built to accurately describe and predict skin colour appearance, e.g. using the perceived skin whiteness, redness, and yellowness instead of the L^*, a^*, and b^* as the new coordinates. With the optimised colour appearance model, it might be possible to map the facia colour cues to a new colour space and achieve better prediction of facial preference judgements.

Lighting can be a factor that largely changes the colour appearance of human faces and thus affect the preference assessment. The present experimental works used facial images from the LLSD to assess the facial colour appearance and preference. The facial images from that database were all captured under the CIE standard illumination D65. In the future, the influence of lighting conditions could be considered and facial colour preference under various lightings could be further investigated. Currently, the colour preference index, especially for light-emitting diodes (LEDs), is

under discussion by CIE technical committee TC1-91, and skin colour is still considered as a special target. A better metric is needed to quantify light source colour rendering and preference properties for human skin colour.

The present research has provided a useful and repeatable methodology for facial colour preference research based on a realistic skin model. Large cultural differences in facial colour perceptions were found between Caucasian and Chinese populations. The future study could make the methodology more inclusive by involving more different ethnic groups. A larger image database and larger-scale psychophysical experiments will be needed to collect more observation data. Both the observer and the observed will cover more different ethnic groups and people from different cultural backgrounds. The cultural difference will be fully discussed by comparing the modelling results of different groups.

The effect of ethnicity has been investigated in our experiments using observers of both genders, the observer gender is another possible additional variable. While the current sample is too small to conduct a meaningful gender-based analysis, previous studies found a strong effect of facial redness that impacts on perceived health and attractiveness for both male and female skin by skin colour manipulations (Pazda et al., 2016; Thorstenson et al., 2017). Whether there is a perceptual difference for facial colour appearance between gender and, if there is, how large the effect is in the realistic skin model compared to the cultural difference, requires further work.

Human colour perception of facial appearance is rich and complicated. In this study, factors that influence the perception of overall facial colour appearance were discussed and three possible assumptions were brought up explaining why such global facial colour impressions formed. Those assumptions will need to be tested. Besides, the perceptual difference in facial colour appearance between Caucasian and Chinese observers was also observed. Further research is still needed to explore whether the facial appearance perception is culture-dependent and whether perceptual differences have come from their natural chromatic properties, state of adaption, sensitivity or other aspects. Additionally, the mechanisms behind the perceptual differences between the face stimuli and the patch stimuli are still unknown. A large amount of research work is required to fully understand the visual sensitivity and the physiological mechanisms underlying facial colour perception.

List of References

Abdi, H. and Williams, L.J. 2010. Principal component analysis. *WIREs Computational Statistics*. **2**(4), pp.433–459.

Abdul-Wahab, S.A., Bakheit, C.S. and Al-Alawi, S.M. 2005. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling and Software*. **20**(10), pp.1263–1271.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*. **19**(6), pp.716–723.

Albers, J. 2013. *Interaction of Colour (1962)*. New Haven and London: Yale University Press.

Amano, K., Xiao, K., Wuerger, S. and Meyer, G. 2020. A colorimetric comparison of sunless with natural skin tan A. Fukunaga, ed. *PLOS ONE*. **15**(12), p.e0233816.

Ansiau, D., Marquié, J.C., Soubelet, A. and Ramos, S. 2005. Relationships between cognitive characteristics of the job, age, and cognitive efficiency. *International Congress Series*. **1280**, pp.43–48.

Appleton, K.M., McGrath, A.J., McKinley, M.C., Draffin, C.R., Hamill, L.L., Young, I.S. and Woodside, J. V 2018. The value of facial attractiveness for encouraging fruit and vegetable consumption: analyses from a randomized controlled trial. *BMC Public Health*. **18**(1), p.298.

ASTM, D. 1993. E313-73: Standard test method for indices of whiteness and yellowness of near-white. *Opaque Mater*.

Balasubramanian, R. 2017. Device characterization *In*: *Digital color imaging handbook*. CRC Press, pp.269–384.

Bates, D., Mächler, M., Bolker, B. and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. **67**(1).

Berns, R.S. 2019. *Billmeyer and Saltzman's Principles of Color Technology*. Wiley.

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

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

Berns, Roy S, Gorzynski, M. El and Motta, R.J. 1993. CRT colorimetry. Part II: Metrology. *Color Research & Application*. **18**(5), pp.315–325.

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

Bilal, S., Akmeliawati, R., Salami, M.J.E. and Shafie, A.A. 2015. Dynamic approach for real-time skin detection. *Journal of Real-Time Image Processing*. **10**(2), pp.371–385.

Del Bino, S. and Bernerd, F. 2013. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *British Journal of Dermatology.* **169**(SUPPL. 3), pp.33–40.

Burnham, K.P. and Anderson, D.R. 2004. Multimodel inference. *Sociological Methods & Research.* **33**(2), pp.261–304.

Burns, P.D. and Berns, R.S. 1996. Analysis multispectral image capture. *Color and Imaging Conference.* **4**(1), pp.19–22.

Burns, S.A., Elsner, A.E., Pokorny, J. and Smith, V.C. 1984. The Abney effect: chromaticity coordinates of unique and other constant hues. *Vision Research.* **24**(5), pp.479–489.

Cai, Z., Hahn, A.C., Zhang, W., Holzleitner, I.J., Lee, A.J., DeBruine, L.M. and Jones, B.C. 2019. No evidence that facial attractiveness, femininity, averageness, or coloration are cues to susceptibility to infectious illnesses in a university sample of young adult women. *Evolution and Human Behavior.* **40**(2), pp.156–159.

Çamdevýren, H., Demýr, N., Kanik, A. and Keskýn, S. 2005. Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling.* **181**(4), pp.581–589.

Chan, S.L. and Park, M. 2005. Project cost estimation using principal component regression. *Construction Management and Economics.* **23**(3), pp.295–304.

Changizi, M.A., Zhang, Q. and Shimojo, S. 2006. Bare skin, blood and the evolution of primate colour vision. *Biology Letters.* **2**(2), pp.217–221.

Chardon, A., Cretois, I. and Hourseau, C. 1991. Skin colour typology and suntanning pathways. *International Journal of Cosmetic Science.* **13**(4), pp.191–208.

CIE 2018. CIE 015: 2018: Colorimetry.

CIE 2022. CIE 248: 2022: The CIE 2016 Colour Appearance Model For Colour Management Systems: CIECAM16.

Clark, J.H. 1924. The Ishihara test for color blindness. *American Journal of Physiological Optics.* (5), pp.269–276.

Clarke, F.J.J., 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.

Coetzee, V., Greeff, J.M., Stephen, I.D. and Perrett, D.I. 2014. Cross-cultural agreement in facial attractiveness preferences: the role of ethnicity and gender C. McCormick, ed. *PLoS ONE.* **9**(7), p.e99629.

Craik, K.J.W. 1939. The effect of adaptation upon visual acuity. *British Journal of Psychology.* **29**(3), p.252.

Cronbach, L.J. 1951. Coefficient alpha and the internal structure of tests. *psychometrika.* **16**(3), pp.297–334.

Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B. and Wu, C.-H. 1995. 'Their ideas of beauty are, on the whole, the same as ours': Consistency and variability in the cross-cultural perception of female

physical attractiveness. *Journal of Personality and Social Psychology*. **68**(2), pp.261–279.

Day, E.A., Taplin, L. and Berns, R.S. 2004. Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display. *Col Res Appl*. **29**, pp.365–373.

Dessinioti, C. and Antoniou, C. 2017. The " red face ": Not always rosacea. *Clinics in Dermatology*. **35**(2), pp.201–206.

Elliot, A.J., Fairchild, M.D. and Franklin, A. 2015. *Handbook of color psychology*. Cambridge University Press.

Fairchild, M.D. 1997. CIE TC1-34 Testing Colour Appearance Models. , p.21.

Fairchild, M.D. 2013. *Color appearance models*. John Wiley & Sons.

Fairchild, M.D. 1995. Testing colour-appearance models: Guidelines for coordinated research. *Color Research & Application*. **20**(4), pp.262–267.

Filzmoser, P. and Nordhausen, K. 2021. Robust linear regression for high-dimensional data: An overview. *WIREs Computational Statistics*. **13**(4).

Fink, B., Bunse, L., Matts, P.J. and D'Emiliano, D. 2012. Visible skin colouration predicts perception of male facial age, health and attractiveness. *International Journal of Cosmetic Science*. **34**(4), pp.307–310.

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. 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.J., D'Emiliano, D., Bunse, L., Weege, B. and Röder, S. 2011. 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.

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

Foo, Y.Z., Rhodes, G. and Simmons, L.W. 2017. The carotenoid beta-carotene enhances facial color, attractiveness and perceived health, but not actual health, in humans. *Behavioral Ecology*. **28**(2), pp.570–578.

Foo, Y.Z., Simmons, L.W. and Rhodes, G. 2017. Predictors of facial attractiveness and health in humans. *Scientific Reports*. **7**(February), pp.1–12.

Gao, Y., Niddam, J., Noel, W., Hersant, B. and Meningaud, J.P. 2018. Comparison of aesthetic facial criteria between Caucasian and East Asian female populations: An esthetic surgeon's perspective. *Asian Journal of Surgery*. **41**(1), pp.4–11.

Gegenfurtner, K.R. 2003. Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*. **4**(7), pp.563–572.

Geladi, P. and Kowalski, B.R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*. **185**, pp.1–17.

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

Giesel, M. and Gegenfurtner, K.R. 2010. Color appearance of real objects varying in material, hue, and shape. *Journal of Vision*. **10**(9), pp.10–10.

Gomes, A.I., Pires, J.C.M., Figueiredo, S.A. and Boaventura, R.A.R. 2014. Multiple linear and principal component regressions for modelling ecotoxicity bioassay response. *Environmental Technology (United Kingdom)*. **35**(8), pp.945–955.

Granzier, J.J.M. and Gegenfurtner, K.R. 2012. Effects of memory colour on colour constancy for unknown coloured objects. *i-Perception*. **3**(3), pp.190–215.

Guild, J. 1931. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*. **230**(681–693), pp.149–187.

Han, C., Wang, H., Hahn, A.C., Fisher, C.I., Kandrik, M., Fasolt, V., Morrison, D.K., Lee, A.J., Holzleitner, I.J., DeBruine, L.M. and Jones, B.C. 2018. Cultural differences in preferences for facial coloration. *Evolution and Human Behavior*. **39**(2), pp.154–159.

Hasantash, M., Lafer-Sousa, R., Afraz, A. and Conway, B.R. 2019. Paradoxical impact of memory on color appearance of faces. *Nature Communications*. **10**(1), pp.1–10.

Hastie, T., Friedman, J. and Tibshirani, R. 2001. *The Elements of Statistical Learning*. New York, NY: Springer New York.

He, Y., Mikami, T., Tanaka, S., Kikuchi, K. and Mizokami, Y. 2021. Comparisons in perception of facial skin brightness, as influenced by differences in skin color: Asian observers. *Color Research & Application*. **46**(4), pp.808–820.

Hering, E. 1964. *Outlines of a Theory of the Light Sense.* trans. Hur. Harvard University Press.

Hoerl, A.E. and Kennard, R.W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. **12**(1), pp.55–67.

Holzleitner, I.J., Lee, A.J., Hahn, A.C., Kandrik, M., Bovet, J., Renoult, J.P., Simmons, D., Garrod, O., DeBruine, L.M. and Jones, B.C. 2019. Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance*. **45**(12), pp.1589–1595.

Howitt, D. and Cramer, D. 2007. *Introduction to statistics in psychology*. Pearson education.

Huixia, Q., Xiaohui, L., Chengda, Y., Yanlu, Z., Senee, J., Laurent, A., Bazin, R., Flament, F., Adam, A. and Piot, B. 2012. Instrumental and clinical

studies of the facial skin tone and pigmentation of Shanghaiese women. Changes induced by age and a cosmetic whitening product. *International Journal of Cosmetic Science*. **34**(1), pp.49–54.

Hunt, R.W.G. 1952. Light and dark adaptation and the perception of color. *JOSA*. **42**(3), pp.190–199.

Hunt, R.W.G. and Pointer, M.R. 2011. *Measuring Colour* 4th ed. Chichester, UK: Wiley.

Igarashi, T., Nishino, K. and Nayar, S.K. 2007. The appearance of human skin: a survey. *Foundations and Trends® in Computer Graphics and Vision*. **3**(1), pp.1–95.

Imai, F.H., Tsumura, N., Haneishi, H. and Miyake, Y. 1996. Principal component analysis of skin color and its application to colorimetric color reproduction on CRT display and hardcopy. *Journal of Imaging Science and Technology*. **40**(5), pp.422–430.

ISO/CIE 2022. ISO/CIE 11664-2:2022 | EN ISO/CIE 11664-2:2022, CIE Colorimetry-Part 2: Standard Illuminants for Colorimetry.

Jaeger, B. and Jones, A.L. 2022. Which Facial Features Are Central in Impression Formation? *Social Psychological and Personality Science*. **13**(2), pp.553–561.

Jaeger, B. and Meral, E.O. 2022. Who Can Be Fooled? Modeling Facial Impressions of Gullibility. *Social Cognition*. **40**(2), pp.127–149.

Janik, S.W., Wellens, A.R., Goldberg, M.L. and Dell'osso, L.F. 1978. Eyes as the center of focus in the visual examination of human faces. *Perceptual and Motor Skills*. **47**(3 Pt 1), pp.857–858.

Johnson, T. 1996. Colour management in graphic art and publication. *Leatherhead: Pira International*.

Jones, A.L. 2018. The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior*. **39**(1), pp.19–29.

Jones, A.L. and Jaeger, B. 2019. Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*. **11**(2).

Jones, A.L., Porcheron, A., Sweda, J.R., Morizot, F. and Russell, R. 2016. Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image*. **17**, pp.57–66.

Jones, A.L., Russell, R. and Ward, R. 2015. Cosmetics alter biologically-based factors of beauty: evidence from facial contrast. *Evolutionary Psychology*. **13**(1), p.147470491501300.

Kikuchi, K., Mizokami, Y., Egawa, M. and Yaguchi, H. 2020. Development of an image evaluation method for skin color distribution in facial images and its application: Aging effects and seasonal changes of facial color distribution. *Color Research & Application*. **45**(2), pp.290–302.

Kimura, E. 2018. Averaging colors of multicolor mosaics. *Journal of the Optical Society of America A*. **35**(4), p.B43.

Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of statistical software*. **28**, pp.1–26.

Kuo, W. 2007. The feasibility of establishing new color image scales using the magnitude estimation method. *Color Research & Application*. **32**(6), pp.463–468.

Kuriki, I. 2004. Testing the possibility of average-color perception from multi-colored patterns. *Optical Review*. **11**, pp.249–257.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software*. **82**(13), pp.1–26.

Lefevre, C.E. and Perrett, D.I. 2015. Fruit over sunbed: Carotenoid skin colouration is found more attractive than melanin colouration. *Quarterly Journal of Experimental Psychology*. **68**(2), pp.284–293.

Little, A.C., Jones, B.C. and DeBruine, L.M. 2011. Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **366**(1571), pp.1638–1659.

Luo, M.R., Clarke, A.A., Rhodes, P.A., Schappo, A., Scrivener, S.A.R. and Tait, C.J. 1991. Quantifying colour appearance. Part I. Lutchi colour appearance data. *Color Research & Application*. **16**(3), pp.166–180.

Luo, M.R., Gao, X.W. and Scrivener, S.A.R. 1995. Quantifying colour appearance. part V. simultaneous contrast. *Color Research & Application*. **20**(1), pp.18–28.

Luo, M.R. and Hunt, R.W.G. 1998. Testing colour appearance models using corresponding-colour and magnitude-estimation data sets. *Color Research & Application*. **23**(3), pp.147–153.

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., Pretty, I. and Cheung, V. 2009. Development of a whiteness index for dentistry. *Journal of Dentistry*. **37**, pp.e21–e26.

Marszalec, E.A., Martinkauppi, J.B., Soriano, M.N. and Pietikaeinen, M. 2000. Physics-based face database for color research. *Journal of Electronic Imaging*. **9**(1), pp.32–38.

Matts, P.J., Fink, B., Grammer, K. and Burquest, M. 2007. Color homogeneity and visual perception of age, health, and attractiveness of female facial skin. *Journal of the American Academy of Dermatology*. **57**(6), pp.977–984.

Mayes, A.E., Murray, P.G., Gunn, D.A., Tomlin, C.C., Catt, S.D., Wen, Y.B., Zhou, L.P., Wang, H.Q., Catt, M. and Granger, S.P. 2010. Ageing appearance in China: Biophysical profile of facial skin and its relationship to perceived age. *Journal of the European Academy of Dermatology and Venereology*. **24**(3), pp.341–348.

McAdams, H.T., Crawford, R.W. and Hadder, G.R. 2000. A vector approach to regression analysis and its application to heavy-duty diesel emissions

In: *SAE transactions*. JSTOR, pp.1735–1752.

McDonald, G.C. 2009. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*. **1**(1), pp.93–100.

McLaren, K. 1986. *The Colour Science of Dyes and Pigments*. Taylor & Francis.

Melgosa, M., Richard, N., Fernández-Maloigne, C., Xiao, K., de Clermont-Gallerande, H., Jost-Boissard, S. and Okajima, K. 2018. Colour differences in Caucasian and Oriental women's faces illuminated by white light-emitting diode sources. *International Journal of Cosmetic Science*. **40**(3), pp.244–255.

Michelson, A.A. 1995. *Studies in optics*. Courier Corporation.

Miyamoto, K., Takiwaki, H., Hillebrand, G.G. and Arase, S. 2002. Development of a digital imaging system for objective measurement of hyperpigmented spots on the face. *Skin Research and Technology*. **8**(4), pp.227–235.

Mountains, B. 2013. Principal component regression analysis in water demand forecasting : an application to the Blue Mountains, NSW, Australia. *Environmental Science*. **1**(1), pp.49–59.

Nadal, M.E., Cameron Miller, C. and Fairman, H.S. 2011. Statistical methods for analyzing color difference distributions. *Color Research & Application*. **36**(3), pp.160–168.

Newton, I. and Hemming, G.. 1704. *Opticks: or, A treatise of the reflections, refractions, inflexions and colours of light : also two treatises of the species and magnitude of curvilinear figures*. London: Printed for Sam. Smith, and Benj. Walford.

Nkengne, A., Bertin, C., Stamatas, G.N., Giron, A., Rossi, A., Issachar, N. and Fertil, B. 2008. Influence of facial skin attributes on the perceived age of Caucasian women. *Journal of the European Academy of Dermatology and Venereology*. **22**(8), pp.982–991.

Ohno, Y. 2000. CIE fundamentals for color measurements *In*: *Digital Printing Technologies, IS&T's NIP16, International Conference, IS&T*. Vancouver, CA.

Ohta, N. and Robertson, A. 2006. *Colorimetry: fundamentals and applications*. John Wiley & Sons.

Okuda, S. and Okajima, K. 2017. Effects of spectral component of light on appearance of skin of woman's face with make-up. *Journal of Light & Visual Environment*. **40**, pp.20–27.

Pazda, A.D., Thorstenson, C.A., Elliot, A.J. and Perrett, D.I. 2016. Women's facial redness increases their perceived attractiveness: mediation through perceived healthiness. *Perception*. **45**(7), pp.739–754.

Pérez, M. del M., Ghinea, R., Rivas, M.J., Yebra, A., Ionescu, A.M., Paravina, R.D. and Herrera, L.J. 2016. Development of a customized whiteness index for dentistry based on CIELAB color space. *Dental Materials*. **32**(3), pp.461–467.

Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M. and

Pereira, M.C. 2008. Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling and Software*. **23**(1), pp.50–55.

Pointer, M.R., Rhodes, P.A., Choi, S. and Luo, M. 2008. Investigation of large display color image appearance I: Important factors affecting perceived quality. *Journal of Imaging Science and Technology*. **52**(4), pp.40904-1-40904–11.

Porcheron, A., Mauger, E. and Russell, R. 2013. Aspects of facial contrast decrease with age and are cues for age perception. *PLoS ONE*. **8**(3).

Porcheron, A., Mauger, E., Soppelsa, F., Liu, Y., Ge, L., Pascalis, O., Russell, R. and Morizot, F. 2017. Facial contrast is a cross-cultural cue for perceiving age. *Frontiers in Physiology*. **8**(JUL).

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.

Purdy, D.M. 1931. Spectral hue as a function of intensity. *The American Journal of Psychology*. **43**(4), p.541.

Ransom, C.J., Kitchen, N.R., Camberato, J.J., Carter, P.R., Ferguson, R.B., Fernández, F.G., Franzen, D.W., Laboski, C.A.M., Myers, D.B., Nafziger, E.D., Sawyer, J.E. and Shanahan, J.F. 2019. Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. *Computers and Electronics in Agriculture*. **164**(March), p.104872.

RDC, T. 2010. A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing*.

Re, D.E., Whitehead, R.D., Xiao, D. and Perrett, D.I. 2011. Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. *PloS one*. **6**(3), p.e17859.

Rhodes, G. 2006. The evolutionary psychology of facial beauty. *Annual Review of Psychology*. **57**(1), pp.199–226.

De Rigal, J., Abella, M., 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.

Rowland, H.M. and Burriss, R.P. 2017. Human colour in mate choice and competition. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372**(1724), p.20160350.

Russell, R. 2009. A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*. **38**(8), pp.1211–1219.

Russell, R. 2003. Sex, beauty, and the relative luminance of facial features. *Perception*. **32**(9), pp.1093–1107.

Russell, R., Batres, C., Courrèges, S., Kaminski, G., Soppelsa, F., Morizot, F. and Porcheron, A. 2019. Differential effects of makeup on perceived age. *British Journal of Psychology*. **110**(1), pp.87–100.

Russell, R., Porcheron, A., Sweda, J.R., Jones, A.L., Mauger, E. and Morizot, F. 2016. Facial contrast is a cue for perceiving health from the

face. *Journal of Experimental Psychology: Human Perception and Performance*. **42**(9), pp.1354–1362.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*. **6**(2), pp.461–464.

Scott, I.M.L., Pound, N., Stephen, I.D., Clark, A.P. and Penton-Voak, I.S. 2010. Does masculinity matter? The contribution of masculine face shape to male attractiveness in humans. *PLoS ONE*. **5**(10).

Sharma, G., Wu, W. and Dalal, E.N. 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*. **30**(1), pp.21–30.

Shimakura, H. and Sakata, K. 2019. Desaturation-induced brightness in face color perception. *i-Perception*. **10**(3), p.204166951985478.

Sohaib, A., Amano, K., Xiao, K., Yates, J.M., Whitford, C. and Wuerger, S. 2018. Colour quality of facial prostheses in additive manufacturing. *The International Journal of Advanced Manufacturing Technology*. **96**(1–4), pp.881–894.

Stephen, I.D., Coetzee, V., Law Smith, M. and Perrett, D.I. 2009. Skin blood perfusion and oxygenation colour affect perceived human health R. Sear, ed. *PLoS ONE*. **4**(4), p.e5083.

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., Law Smith, M.J., Stirrat, M.R. and Perrett, D.I. 2009. Facial skin coloration affects perceived health of human faces. *International Journal of Primatology*. **30**(6), pp.845–857.

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.L., 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.

Stevens, J.C. and Stevens, S.S. 1963. Brightness function: Effects of adaptation. *JOSA*. **53**(3), pp.375–385.

Stevens, S.S. 1946. On the theory of scales of measurement. *Science*. **103**(2684), pp.677–680.

Sullivan, C., Pan, Q., Westland, S. and Ellwood, R. 2019. A yellowness index for use in dentistry. *Journal of Dentistry*. **91**, p.103244.

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.

Sunaga, S. and Yamashita, Y. 2007. Global color impressions of multicolored textured patterns with equal unique hue elements. *Color Research and Application*. **32**(4), pp.267–277.

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., Haneishi, H., Ojima, N. and Tsukada, M. 2002. Representative data selection for standard object colour spectra database (SOCS). *Color and Imaging Conference*. **10**(1), pp.155–160.

Tajima, J., Tsukada, M., Miyake, Y., Haneishi, H., Tsumura, N., Nakajima, M., Azuma, Y., Iga, T., Inui, M., Ohta, N., Ojima, N. and Sanada, S. 1998. Development and standardization of a spectral characteristics database for evaluating color reproduction in image input devices *In*: J. Bares, ed. *Electronic Imaging: Processing, Printing, and Publishing in color*. SPIE, pp.42–50.

Tan, K.W., Graf, B.A., Mitra, S.R. and Stephen, I.D. 2015. Daily Consumption of a Fruit and Vegetable Smoothie Alters Facial Skin Color P. V. Nerurkar, ed. *PLOS ONE*. **10**(7), p.e0133445.

Tan, K.W. and Stephen, I.D. 2013. Colour detection thresholds in faces and colour patches. *Perception*. **42**(7), pp.733–741.

Tan, K.W. and Stephen, I.D. 2019. Skin Color Preferences in a Malaysian Chinese Population. *Frontiers in Psychology*. **10**(JUN).

Tan, K.W., Tiddeman, B. and Stephen, I.D. 2018. Skin texture and colour predict perceived health in Asian faces. *Evolution and Human Behavior*. **39**(3), pp.320–335.

Taylor, S.C., Arsonnaud, S., Czernielewski, J. and Hyperpigmentation Scale Study Group 2005. The Taylor Hyperpigmentation Scale: a new visual assessment tool for the evaluation of skin color and pigmentation. *Cutis*. **76**(4), pp.270–4.

Teghtsoonian, R., Stevens, S.S. and Stevens, G. 1975. Psychophysics: Introduction to its perceptual, neural, and social prospects. *The American Journal of Psychology*. **88**(4), p.677.

Thornhill, R. and Gangestad, S.W. 1999. Facial attractiveness. *Trends in Cognitive Sciences*. **3**(12), pp.452–460.

Thorstenson, C.A., Pazda, A.D. and Elliot, A.J. 2020. Social perception of facial color appearance for human trichromatic versus dichromatic color vision. *Personality and Social Psychology Bulletin*. **46**(1), pp.51–63.

Thorstenson, C.A., Pazda, A.D., Elliot, A.J. and Perrett, D.I. 2017. Facial redness increases men's perceived healthiness and attractiveness. *Perception*. **46**(6), pp.650–664.

Thurstone, L.L. 1927. A law of comparative judgment. *Psychological Review*. **34**(4), pp.273–286.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. **58**(1), pp.267–288.

Torgerson, W.S. 1958. *Theory and methods of scaling*. New York: John Wiley and Sons, Inc.

Virtanen, L.S., Olkkonen, M. and Saarela, T.P. 2020. Color ensembles: Sampling and averaging spatial hue distributions. *Journal of Vision.* **20**(5), pp.1–14.

Vos, J.J. and Walraven, P.L. 1971. On the derivation of the foveal receptor primaries. *Vision Research.* **11**(8), pp.799–818.

Wandell, B.A. 1995. *Foundations of vision.* Sinauer Associates.

Wang, M., Xiao, K., Luo, M.R., Pointer, M., Cheung, V. and Wuerger, S. 2018. An investigation into the variability of skin colour measurements. *Color Research & Application.* **43**(4), pp.458–470.

Webster, M.A., Kaping, D., Mizokami, Y. and Duhamel, P. 2004. Adaptation to natural facial categories. *Nature.* **428**(6982), pp.557–561.

Webster, M.A. and MacLeod, D.I.A. 2011. Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences.* **366**(1571), pp.1702–1725.

Westland, S., Ripamonti, C. and Cheung, V. 2012. *Computational colour science using MATLAB.* John Wiley & Sons.

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), p.147470491201000.

Wright, W.D. 1929. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society.* **30**(4), p.141.

Wyszecki, G. 1967. Correlate for lightness in terms of CIE chromaticity coordinates and luminous reflectance. *JOSA.* **57**(2), pp.254–257.

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), p.83301.

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), pp.21–29.

Xiao, K., Zardawi, F., van Noort, R. and Yates, J.M. 2013. Color reproduction for advanced manufacture of soft tissue prostheses. *Journal of Dentistry.* **41**, pp.e15–e23.

Xiao, K., Zardawi, F., van Noort, R. and Yates, J.M. 2014. Developing a 3D colour image reproduction system for additive manufacturing of facial prostheses. *The International Journal of Advanced Manufacturing Technology.* **70**(9–12), pp.2043–2049.

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.

Xie, Q. (Vivi) and Zhang, M. 2013. White or tan? A cross-cultural analysis of skin beauty advertisements between China and the United States. *Asian Journal of Communication.* **23**(5), pp.538–554.

Yamashita, T., Yamashita, K. and Kamimura, R. 2007. A stepwise AIC method for variable selection in linear regression. *Communications in Statistics - Theory and Methods*. **36**(13), pp.2395–2403.

Yoshikawa, H., Kikuchi, K., Yaguchi, H., Mizokami, Y. and Takata, S. 2012. Effect of chromatic components on facial skin whiteness. *Color Research & Application*. **37**(4), pp.281–291.

Young, T. 1845. *Young, Thomas. A Course of Lectures on Natural Philosophy and the Mechanical Arts (Miscellaneous Papers, Reprinted with Corrections)*. London: Taylor and Walton.

Zeng, H. and Luo, R. 2010. Modelling memory colour region for preference colour reproduction *In*: R. Eschbach, G. G. Marcu, S. Tominaga and A. Rizzi, eds. *Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*. International Society for Optics and Photonics, p.752808.

Zou, H. and Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. **67**(2), pp.301–320.

List of Abbreviations

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CA	Caucasian
CI	Confidence Interval
CIE	Commission Internationale de L' Éclairage (International Commission on Illumination)
CAM	Colour Appearance Model
CIELAB	CIE 1976 (L∗a∗b∗) Colour Space
CN	Chinese
CRT	Cathode Ray Tube
CV	Coefficient of Variation
DV	Dependent Variables
EN	The Elastic Net Regression
FV	Fruit and Vegetable
GOG	Gamma Offset Gain
ISO	International Organization for Standardization
ITA	The Individual Typology Angle
IV	Independent Variable
LASSO	Least Absolute Shrinkage and Selection Operator Regression
LCD	Liquid Crystal Display
LED	Light-Emitting Diodes
LGN	lateral geniculate nucleus
LLSD	Liverpool-Leeds Skin-colour Database
MCDM	The Mean Colour Difference from The Mean
MLR	Multiple Linear Regression
NS	Not Significant
OLS	Ordinary Least Squares Regression
PCA	Principal Component Analysis
PCR	Principal Component Regression

PLCC	The Piecewise Linear Assuming Chromaticity Constancy Model
PLSR	Partial Least Squares Regression
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
RIS	Redness Index for Skin
RMSE	Root Mean Squared Error
RR	Ridge Regression
SB	Stepwise Regression - Backward Steps
SD	Standard Deviation
SF	Stepwise Regression – Forward Steps
SLR	Single Lens Reflex
SOCS	Standard Object Colour Spectra Database
SP	Spectrophotometers
TC	Technical Committee
TSR	Tele-spectroradiometers
UV	Ultra Violet
WIC	CIE Whiteness Index
WIS	Whiteness Index for Skin
YIS	Yellowness Index for Skin