Prototypes of facial attributes developed through image averaging techniques

S.S. HAWKINS1*, D.I. PERRETT2, D.M. BURT2, D.A. ROWLAND2 and R.I. MURAHATA1

1Unilever Research US, Edgewater, NJ, USA; 2School of Psychology, University of St. Andrews, Fife, KY16 9JU, UK

Poster Presentation of the 20th Congress, International Federation of Societies of Cosmetic Chemists, Cannes, France, September 14–18, 1998

Received 26th May 1998
Accepted 1st November 1998

Keywords: facial averaging; caricaturing; healthy glow

Synopsis
Image capture and quantification has proven useful in a variety of scientific applications, for example, biology, medicine, geology, meteorology and forensics. The objective of this research was to utilize this technology to quantify clinical- and consumer-perceivable changes in facial attributes. A panel of expert assessors was trained, and, in a large consumer study, consumer facial attributes were identified and grading scales for each attribute were established. These experts then rated over 240 subjects on a total of 19 different facial attributes. Based on methodology developed by Perrett et al., facial averages or prototypes were computed from panelists rated high or low for each attribute. Prototypes were developed in a 3 step process: 1) selection of 224 predefined feature points; 2) calculation of average face shape; and 3) 'morphing' individual faces into that shape and blending the images together. Naive assessors could readily appreciate the differences in facial appearance of the prototypes. In addition, expert graders were able to identify the general class of attribute affected. This method provides a powerful tool for assessing the effects of skin care technologies.

* To whom correspondence should be addressed.

0142–5463 © 1999 International Journal of Cosmetic Science
Introduction

The term ‘healthy skin’ is an important contributor to consumer perceptions of ideal skin for face and body. Further, healthy skin for the face is a major unsatisfied consumer need. It is important to note that this consumer language is concerned with a cosmetic or appearance-based definition of health, not a medical one. Previous qualitative and quantitative work provided a verbal framework and identification of individual descriptors that contributed to overall healthy appearance. An expert panel of healthy skin evaluators was used as the first step in quantifying healthy skin attributes. It was desirable, however, to develop an instrumental, highly objective technique for studying these attributes.

A number of papers describe an ideal relationship of vertical and other structural proportions on the face, and how these structures relate to the appearance of facial health. However, the approach taken was on true medical conditions rather than cosmetic appearance per se. The authors also provide predictions for changes/improvements in facial appearance that are related to changes in underlying structures (soft tissue, bone), based on orthodontic case studies or observations of computer-altered images of the face. These predictions are based on morphological studies in medicine, and are based on underlying structural changes rather than cosmetic-based treatment and change [1–3].

Many other researchers in psychology and perception have studied the apparent improvement of facial beauty using averaging techniques, or an improvement in recognition based on facial caricaturing [4–8].

The previous work has not been applied specifically to model individual healthy attributes, which taken collectively define a composite cosmetic attribute (e.g., ‘healthy skin’), sought by a consumer. To date, there have been reports demonstrating apparent changes in perceived age or gender through alteration of single or a small cluster of attributes. The relationship of features of the face in the perception of age and gender has also been studied, but not for the development of the expert composite cosmetic attribute (e.g., ‘healthy skin’) ratings and prototypes discussed here [9–13]. None have described changes in complex composite attributes such as healthy nor have any been specifically related to evaluations by or the development of an expert panel rating methodology.

Methods

‘Healthy skin’ is a complex composite cosmetic attribute, that has not been previously well defined. Qualitative language generation with Caucasian females in the US was performed to determine the individual attributes of healthy looking facial skin. Various cosmetic facial attributes, such as lines and wrinkles, pores, blemishes, sheen, glow/color, were identified as consumer perceivable and consumer defined attributes that contribute to the overall appearance of cosmetic health. This language was subsequently validated with other fair skinned populations in other parts of the world. Using standard techniques expert graders were trained to evaluate a reduced set of these descriptors.

The results of two studies conducted to provide quantitative image information on healthy skin characteristics are reported here:
Facial prototypes through image averaging techniques

Study #1

A pilot study was conducted with 80 female subjects judged by the expert assessors to have a range of overall skin health. The objective of this study was to determine the feasibility of generating healthy attribute facial prototypes and caricatures using the methodology described below. The subjects all wore white hair caps and black bibs. Photographs were recorded using a Nikon F3 35mm camera.

Study #2

A large-scale study with 240 female subjects was conducted to validate the original attribute prototypes and caricatures. Rather than wearing white hair caps and black bibs, subjects were instructed to move their hair away from their face as much as possible. Images were recorded with a Sony DCR-VX1000 digital camera under varying lighting conditions, to find the dependency on capture techniques for generating prototypes and caricatures.

In both studies, expert evaluations and photographs were taken of subjects under ‘baseline’ conditions (normal habit of treatment/product use). Subjects cleansed their faces with their normal cleansing products approximately 2 hours prior to their arrival at the study site. Photographs were recorded to PhotoCDs for processing.

Based on methodology developed by Perrett et al. [11] facial averages or prototypes were computed from subjects rated high or low for each individual attribute, including overall healthy skin appearance. Prototypes were developed in a 3 step process:

1) selection of 224 predefined feature points
2) calculation of average face shape
3) ‘morphing’ individual faces into that shape and blending the images together.

Prototypes were computed for a group of subjects with the lowest and highest ratings for healthy glow/color, healthy sheen, and healthy overall appearance. Difference images were than constructed by subtracting the low prototype from the high prototype, and caricatures were generated by adding $2.5 \times$ difference image to the high prototype and subtracting $2.5 \times$ difference image from the low prototype. The methods for computing facial averages and caricatures have been previously described [11].

The resulting caricatures were therefore enhancements of the global differences detected between subjects rated high and low for a particular healthy attribute. This procedure was repeated for a total 19 attributes, that define the composite cosmetic attribute ‘healthy skin.’

Results

Study #1

Naive assessors (University students) could easily distinguish prototypes developed based on overall healthy appearance, and they preferred the relevant prototype for overall health. The relevance of the image manipulation was further documented by presenting the prototypes and caricatures for healthy glow/color, lines and wrinkles, and pores to the expert assessors. The expert assessors were able to identify the correct direction for an attribute (for example positive for healthy glow, negative for blemishes, etc.). In addition, the expert assessors were able to identify the individual attribute that was used in
developing the prototype and caricature images. As the current attributes are not totally independent, it was not surprising that the expert assessors also identified additional attributes that they perceived as significantly changing in the prototypes and caricatures for

Figure 1. Correlation between healthy glow/color facial average and color in the image. Areas of high correlation are shown as gray overlays on the actual color facial average.
healthy glow/color. Blotchiness, blemishes, scars, evenness of color, and dark circles were also perceived as changing in the prototypes and caricatures for high and low healthy glow/color.
Based on the same images, a systematic method using multi-linear regression analysis was developed to identify which areas of the face may be important in the perception of a particular attribute. Those areas in which there is a strong correlation between the healthy

Figure 3. Top left: Prototype for low healthy overall; Top right: Prototype for high healthy overall; Bottom left: Caricature for low healthy overall; Bottom right: Caricature for high healthy overall.
glow/color attribute and the color in the image are overlaid on the facial prototype as gray pixels (see Fig. 1). As can be seen from Fig. 1, ratings of the healthy glow/color are related to color changes in the lateral cheek regions. There is also a relationship between the ratings of healthy glow/color and the color of the regions of skin just below the eyes and nose.

**Study #2**

Results from a larger study confirmed those obtained from Study #1, despite the alterations in capture techniques. Expert assessors were able to identify the correct direction for an attribute (for example positive for healthy glow, negative for blemishes, etc.), as seen in Study #1. The prototypes and caricatures for healthy glow/color and healthy overall appearance are shown in Fig. 2 and 3, respectively. The healthy glow/color prototypes most strongly correlate to contributions in the lateral cheek regions. In addition, there appears to be an age component in the rating of healthy glow/color (the prototype face for high healthy glow/color appears younger).

Since the caricatures enhance the differences between the prototypes, changes between features, contrast in the images, and color changes are dramatically enhanced. In the caricature for high healthy glow (see Fig. 2), for example, the pinkness in the cheeks may not be perceived as healthy, but chapped or irritated. Similarly the high overall healthy caricature in Fig. 3 may not be perceived as healthy due to the perceived overall color. The differences between high and low caricatures, however, can be used to demonstrate the regions and colors of interest for a particular healthy attribute.

**Conclusions**

This report demonstrates the first application of image manipulation technology working with expert assessors and consumer defined complex cosmetic attributes. The ability to quantify the components of a complex consumer concept, such as ‘healthy skin’ on the face provides several advantages. It permits clearer communication with the consumer, and provides the ability to truly target emerging technologies against the appropriate attributes.

**References**


1 Burt and Perrett [9,10] developed analysis procedures based on how well multi-linear regression could model the relationship between RGB (red, green, blue) color of individual pixels in the image and ratings of facial health. Areas of the face for which 10% or more of the variance in ratings could be explained using a linear model are displayed in gray. For pixels representing areas in which the correlation between color and ratings is weaker (< 10% of variance explained) the actual skin color of the average face is displayed.