



Published in final edited form as:

Laryngoscope. 2008 June ; 118(6): 962–974. doi:10.1097/MLG.0b013e31816bf545.

Evolving Attractive Faces Using Morphing Technology and a Genetic Algorithm: A New Approach to Determining Ideal Facial Aesthetics

Brian J. F. Wong, MD, PhD, Koohyar Karmi, Zlatko Devcic, Christine E. McLaren, PhD, and Wen-Pin Chen, MS

Department of Otolaryngology–Head and Neck Surgery (B.J.F.W.), the Department of Biomedical Engineering (B.J.F.W.) the Beckman Laser Institute and Medical Clinic (B.J.F.W., K.K., Z.D.), the Department of Epidemiology (C.E.M.), and the Chao Comprehensive Cancer Center (W.-P.C.), University of California Irvine, Orange, California, U.S.A.

Abstract

Objectives—The objectives of this study were to: 1) determine if a genetic algorithm in combination with morphing software can be used to evolve more attractive faces; and 2) evaluate whether this approach can be used as a tool to define or identify the attributes of the ideal attractive face.

Study Design—Basic research study incorporating focus group evaluations.

Methods—Digital images were acquired of 250 female volunteers (18–25 y). Randomly selected images were used to produce a parent generation (P) of 30 synthetic faces using morphing software. Then, a focus group of 17 trained volunteers (18–25 y) scored each face on an attractiveness scale ranging from 1 (unattractive) to 10 (attractive). A genetic algorithm was used to select 30 new pairs from the parent generation, and these were morphed using software to produce a new first generation (F1) of faces. The F1 faces were scored by the focus group, and the process was repeated for a total of four iterations of the algorithm. The algorithm mimics natural selection by using the attractiveness score as the selection pressure; the more attractive faces are more likely to morph. All five generations (P-F4) were then scored by three focus groups: a) surgeons (n = 12), b) cosmetology students (n = 44), and c) undergraduate students (n = 44). Morphometric measurements were made of 33 specific features on each of the 150 synthetic faces, and correlated with attractiveness scores using univariate and multivariate analysis.

Results—The average facial attractiveness scores increased with each generation and were 3.66 (+0.60), 4.59 (±0.73), 5.50 (±0.62), 6.23 (±0.31), and 6.39 (±0.24) for P and F1–F4 generations, respectively. Histograms of attractiveness score distributions show a significant shift in the skew of each curve toward more attractive faces with each generation. Univariate analysis identified nasal width, eyebrow arch height, and lip thickness as being significantly correlated with attractiveness scores. Multivariate analysis identified a similar collection of morphometric measures. No correlation with more commonly accepted measures such as the length facial thirds or fifths were identified. When images are examined as a montage (by generation), clear distinct trends are identified: oval shaped faces, distinct arched eyebrows, and full lips predominate. Faces evolve to approximate the guidelines suggested by classical canon. F3 and F4 generation faces

look profoundly similar. The statistical and qualitative analysis indicates that the algorithm and methodology succeeds in generating successively more attractive faces.

Conclusions—The use of genetic algorithms in combination with a morphing software and traditional focus-group derived attractiveness scores can be used to evolve attractive synthetic faces. We have demonstrated that the evolution of attractive faces can be mimicked in software. Genetic algorithms and morphing provide a robust alternative to traditional approaches rooted in comparing attractiveness scores with a series of morphometric measurements in human subjects.

Keywords

Facial Beauty; attractiveness; plastic surgery; morphing; genetic algorithm

Introduction

In our culture, being beautiful has its advantages, as we are a society prone to judge a book by its cover. Beautiful people are invested by others with a plethora of desirable characteristics such as warmth, sensitivity, poise, and kindness. Attractive people receive preferential treatment and have intrinsic social, marital, and occupational success as a consequence of winning a genetic lottery. Despite the importance of beauty in our cultural, social, and economic fabric, rigorous definitions of beauty are lacking, and this is particularly true with respect to facial esthetics.¹ Defining beauty remains elusive, though operationally, to paraphrase Supreme Court Justice Potter Stewart: “You know it when you see it.”

Quantitative approaches to defining beauty are rooted in morphometric techniques largely aimed at identifying geometric relationships between facial features and subunits or defining specific linear and angular measurements. Da Vinci and Durer independently developed the classical canons of facial beauty that have permeated art, science, fashion, and popular culture, and set forth the basis for the rules of thirds and fifths and other strategems. Their work has stood the test of time and remains in good agreement with most modern studies on facial proportion. With the rise of mass media through the 20th century, art, popular culture, and fashion converged, and defining facial beauty became relevant to marketing and advertising. The economic impact spurred serious academic inquiry.² With the rise and increasing acceptance of cosmetic surgery during the 1990s, defining beauty has become even more relevant to surgeons.³

Modern approaches combine anthropomorphic methods with focus group ratings of facial beauty.⁴ Focus groups are formed using either expert or lay groups of evaluators who score, rank, or segregate faces based subjectively on appearance. These beauty scores are then tabulated and may be correlated with linear or angular measurements of either the face or photographs of the face taken from different vantage points.⁵

Farkas has done the most detailed and comprehensive work using this basic methodology,⁶ and has published more than 100 articles on this topic alone. His studies are extremely meticulous and involve the use of intricate and innovative devices and techniques for obtaining facial measurements. He has performed studies using widely divergent study subjects across ethnicities, racial groups, and genders, and he has used different types of focus groups as well. By necessity, these comprehensive studies are labor and time intensive, thus limiting the scope and extent of both study subjects and evaluators. Others have adopted his general approach and have teased out cultural influences, segregated focus groups, and further explored demographic influences. While rigorously quantitative, these measures are of limited practical value to the artist, esthetician, marketing executive, or surgeon.

In the 1990s, growing interest in the hypotheses that beauty is rooted in the genetic makeup of the individual and is an indirect measure of overall health, and perhaps more accurately, reproductive fitness, spurred biologists and experimental psychologists to explore this concept in greater detail.⁷ The most celebrated examples of this hypothesis are the studies that examined cross-cultural preferences of men for specific hip-waist-bust ratios in women. The hip-waist-bust ratio is believed to be an indirect link to the subject's hormonal milieu, secondary sexual characteristics, and more broadly, to fertility. In the face, the appeal for men for women with full lips and small jawlines has been hypothesized to correlate with hormonal changes in postadolescent females (again a fertility cue), while in men, a strong jaw and prominent brow ridge are characteristics associated with testosterone surges at maturity.⁸⁻¹² Thus, what we may consider “attractive or beautiful” may be related to structural or functional consequences that are rooted in evolution.^{7,13} These hypothesis rooted in evolutionary biology are speculative, but have been the intriguing subjects of intense academic and popular cultural debate.¹⁴

More recently, digital image processing techniques have been used to alter images and refine which features are found appealing to study populations.¹⁵ The pioneering work of Johnston incorporated custom software with an on-line voting system used to rate faces, and marked a novel approach to identifying the specific appearance of a beautiful face while avoiding the labor intensive approach of traditional morphometric approaches.^{13,16,17} Johnston's landmark body of work identified the “most beautiful face” which was evolved from an expansive on-line voting scheme. Johnston's software “drew” faces based essentially on the outcome of on-line voting, created new faces, collected new votes, and reposted the faces again using an iterative process. While the results of this study are compelling, the software drew faces that provide quite a bit of detail on facial shape and specific features such as eyes and lips; but fell short in terms of producing a realistic digital simulacrum, and was limited because the software made changes in discrete rather than continuous increments. In contrast, others have focused on generating images using advanced image processing or morphing technology, and have examined the impact of specific changes in facial features such as facial shape, and deviation from classical canons in terms of facial proportions, using photographs of real subjects. These studies have required investigators to alter physical features in an ad hoc manner and allowed identification of whether focus groups prefer specific features such as a larger or smaller jaw.

Digital image manipulation in this arena has not been fully exploited and used in defining what is facial beauty, particularly now with the availability of low-cost software and high-powered computing.¹⁸ Currently, digital photographs can be morphed with one another using consumer-level software to produce extremely realistic synthetic faces. Johnston's work is the closest to what may be described as an evolutionary biology approach toward identifying the features of an attractive face, but does not incorporate the randomness inherent in natural selection.

In this study, we used morphing software to create realistic appearing synthetic faces from digital photographs of volunteers. Selection of morphing pairs was accomplished using a genetic algorithm with facial beauty as the only selection pressure. The digital “breeding process” aimed to evolve progressively more attractive facial cohorts with each iteration of the algorithm. The objectives of this study were to: 1) determine if a genetic algorithm in combination with morphing software can be used to evolve more attractive faces; and 2) evaluate whether this approach can be used as a tool to define or identify the attributes of the ideal attractive face.

Materials and Methods

Photography and Subject Population

Digital portraits were taken of women between ages 18 and 25 with the approval of the Institutional Review Board at the University of California Irvine. This study is exclusively focused on female faces; companion studies to follow will examine male faces. No candidates were rejected on the basis of ethnicity or race. Volunteers were rejected if they had obvious craniofacial abnormalities such as cleft lip and cleft palate deformities. Volunteers were solicited from various courses, student associations, sororities, medical student associations, and also from placement of a booth within the University of California Irvine Student Center. A total of 250 volunteers were photographed. Participants were photographed under standard conditions with the face oriented along the Frankfort plane against a neutral blue background. The hair was pulled back with a headband to fully expose the entire face, including the ears and trichial line. A black barber's cape encircled the neck at the level of the sternal notch. At most, only scant natural makeup was permitted, and most subjects were asked to remove their cosmetics and appear clean scrubbed. Only photographs with neutral facial expression (repose) were used. Volunteers were asked to remove earrings and other facial piercings. Digital cameras (Rebel XT, 100 mm Macro Lens, Cannon USA, Lake Success, NY) were used to obtain all images, and faces were photographed at a distance of approximately 6 feet with either flash or ambient artificial lighting.

Morphing Approach

Morphing is the processing of digitally transforming one image into another. Morphing algorithms work by marking prominent features or *registry points*, such as tips and corners, on each of the images. Algorithms are then used to map the movements of these points from one object to the other. The morphing process can be stopped at any point to get different proportions of the first and second image. In this study, we selected Morphman 2000 (STOIK Imaging, LTD, Moscow, Russia) because of its low-cost, ease of use, and capability to use polygonal regions of interest to outline detailed structures such as the eyes, nose, and lips. The software also provided dynamic visualization of both parent images during the registration process. Each synthetic image was a 50:50 morph of two other images.

Through trial and error, we determined that to create highly realistic faces, the pupils, iris, lid crease, eyelashes, vermillion border, eyebrows, alar crease, nasal tip, and ala needed to be identified and outlined with extreme precision on both parent images (Fig. 1). Further polygonal regions of interest over broad featureless regions such as the cheeks, forehead, and chin needed to be encircled, as did the melolabial folds and mental crease. Research assistants constructed preliminary morphing templates initially for each pair of faces. The authors then optimized the templates to improve registration around key features such as the eyes, eyelids, brows, and ears.

Construction of the Parent Generation

Development of the parent generation (P) for morphing presented a logistical challenge. The facial photographs used in this work are part of a larger photograph database managed by the lead author under approval of the Institutional Review Board at the University of California Irvine. This database is being used for several facial analysis projects involving hundreds of subjects. The presentation of actual subject photographs in public venues such as conferences or in publications would require execution of a lengthy written informed consent document. The time required for informed consent would severely limit the accrual of subjects and decrease the number of photographs within our overall database. However, the Institutional Review Board at our institution permitted the use of photographs that have been digitally altered to produce synthetic images such as those created during the morphing

process. Hence, we opted to use synthetic faces for the original parent generation of faces in this study.

The parent generation of faces were produced by first segregating faces into four ethnic groups: 1) white, 2) Asian, 3) Latino, and 4) Middle Eastern. (There were few African-Americans student volunteers in the study, as they make up less than 2% of our county's population.) Photographs within a specific ethnic group were then randomly selected to form pairs and morphed. Thirty pairs of faces were used to generate the parent generation.

Initial Focus Group Evaluations

The parent generation morphs were evaluated and scored from 1 (unattractive) to 10 (attractive) by undergraduate students ($n = 17$) during a one-semester esthetic surgery seminar taught by the lead author over a 12-week period. This small focus group was used to provide facial attractiveness scores because the same students would be available over the full 12-week term. The demographics of this evaluator group reflect the socioeconomic and ethnic composition of undergraduates at our institution and mirror the demographics of our geographic region. Two-thirds of the students were women. Prior to scoring faces, each student evaluator spent a week developing a visual analogue scale for facial beauty with a face (culled from the Internet) representing each score from 1 (unattractive) to 10 (attractive). The use of the visual analogue scale was aimed at encouraging a more consistent approach to scoring faces by each evaluator. The scoring of each face was performed using a classical focus group approach. Images of each face were presented one at a time onto a projection screen using an LCD projector for approximately 45 seconds. Only 30 faces were presented on any given day. Scores for each face were tabulated and averaged, thus providing an average facial attractiveness score for each face in the parent generation. Images for each new generation of evolved morphs were later presented on three additional occasions approximately 3 weeks apart. Of note, this focus group did not evaluate the fourth generation of morphed faces.

Genetic Algorithm

Natural selection is the foundation of biology. It is the process by which favorable traits that are heritable become more common in successive generations, and unfavorable traits become less common. Natural selection acts on the observable characteristics of an organism, favoring individuals with the traits that favor survival and reproduction in a given environment. Over time, this process can result in adaptations that optimize organisms for specific environmental conditions; in humans, evidence of this can be seen in the evolution of different racial groups. Evolving more attractive faces in this study requires the adoption of a heuristic that emulates the process of natural selection. The trait we seek to amplify is facial attractiveness. This cannot be achieved by simply morphing images randomly together as there is no selection pressure. The absence of selection pressure in any combinatorial schema would result in an image with average features. Therefore, we introduced a selection pressure into our algorithm that biased the digital "breeding" process toward selecting more attractive faces.

The basic algorithm is illustrated in Figure 2. First, faces are randomly selected from the parent generation of faces (P). Each face has an attractiveness score associated with it determined by the initial focus group evaluation (see above). Each generation of new faces has a mean, maximum, and minimum attractiveness score, which were produced by the initial evaluation group. In P, the initial focus group produced mean values trending toward a value of 5, and scores close to 1 (profoundly unattractive) or 10 (profoundly attractive) were nonexistent. Second, a random number generator (continuous uniform distribution) returns a value that lies between the minimum and maximum attractiveness score for the

parent generation of faces. Thus, each P face has an attractiveness score and a random number associated with it. Third, each face's attractiveness score is compared to its paired random number. If the attractiveness score exceeds the value produced by the random number generator, then the face can be morphed with another face that also satisfies this condition (i.e., the face is *fit to morph*). Faces selected where the attractiveness score is less than the value produced by the random number generator do not go on to morph, though they may still be selected again later on.

It must be emphasized that “fitness” to morph or *digitally breed* is a function of the attractiveness score and the probability that this score exceeds a random number. Hence, unattractive faces can be selected for morphing if paired with a very low random number, and attractive faces may be rejected if paired with a random number higher than its attractiveness score. But to be sure, the bias is toward the best looking faces. The initial P generation consisted of 30 faces (see above). The algorithm executes until 30 new pairs of faces are generated. Notably, some faces in the original P generation may be represented more than once in this new parent breeding generation (P_b). Likewise, some faces may not be included within any of the 30 pairs in P_b .

The thirty facial pairs in P_b were then morphed to produce the first generation of synthetic faces (F1). F1 faces were then evaluated by the same initial focus group of evaluators, and attractiveness scores were obtained for these new synthetic faces. The genetic algorithm was run, and a subset of F1 faces was selected, namely the breeding cohort $F1_b$. The 60 $F1_b$ faces were then morphed using the genetic algorithm to produce a new second generation of synthetic faces (F2). This process was repeated, producing a third (F3) and fourth (F4) generation. It must be emphasized that while breeding pairs are randomly selected, each face is subject to a selection pressure. The approach mimics the concept of a predator and prey inasmuch as the survival of the prey depends on the fitness of the predator as much as its own. Notably, as in nature, faces are not eliminated from potential “breeding”/morphing after selection and return to the facial “gene pool.”

Average attractiveness scores for each new generation (F1–F3) were calculated via evaluation by the initial focus group. As noted above, 2 to 3 weeks elapsed between the evaluation of each new generation, as that was the time required to produce high quality facial morphs.

Morphometric Measurements

All faces (P-F4) were scaled to the same size using software (Powerpoint, Microsoft, Redmond, WA) with the constraint that distance from the trichial line to the lowest point on the chin (menton) was identical on each image when ported into a Powerpoint slide. This served as a normalization factor. Then each slide of the Powerpoint file was printed using a color laser printer. Thirty-three linear measurements of specific linear features (Table I) on the face were measured. The location of these measurements is noted in the set of diagrams in Figure 3. To increase clarity, some symmetric measurements were not labeled (i.e., only left or right side features were labeled). The measured features are rudimentary and are derived from basic facial proportions described in most plastic surgery textbooks. A particular emphasis has been placed on the eyes and lips as qualitative trends were observed with each generation (see results section). Measurements were obtained using a digital micrometer (Mitutoyo-USA, Aurora, IL), and tabulated for each face.

Evaluation of All Generations Using Additional Focus Groups (Final Focus Groups)

All 150 images for the five generations of morphs were presented in random order to three distinct focus groups for evaluation. The order of the images was randomized. Focus groups

consisted of: 1) undergraduate student volunteers (n = 44); 2) attending surgeons, fellows, and residents in the Department of Otolaryngology-Head and Neck Surgery at XYY (n = 12); and 3) cosmetology school students at a local beauty school (n = 44). The undergraduate students were selected because their age distribution is similar to that of the study subjects, and they are readily accessed through an experimental psychology research participation pool offered by the School of Social Sciences at our institution. The latter two groups were selected as they were thought to have some formal expertise with respect to facial analysis. The undergraduates and the cosmetology students did not know that all of the images were synthetic. The surgeons were aware of image processing, but unaware of the precise details, algorithms, software, or intent of the study. For each of the three groups, images were presented on a projection screen using an LCD projector. In an effort to reduce arbitrary assignment of attractiveness scores, a visual analogue scale was presented before the actual scoring commenced. The visual analogue scale was produced by constructing a montage of faces for each point on the scale from 1 to 10. Source images for the scale were taken from each of the visual analogue scales developed by the original evaluating group described above. The morphed images were then presented, and evaluators recorded their attractiveness scores on a score sheet. There were no incomplete score sheets as each evaluator scored all 150 faces.

Statistical Methods

Univariate analysis was performed for data collected from each of the secondary rater groups (undergraduates, surgeons, and cosmetologists). For each rater group, the average beauty score for each of the 150 faces was computed, the distribution of average beauty scores was examined, and descriptive statistics were computed (median, mean, standard deviation, minimum, and maximum). Similarly, descriptive statistics were computed to examine the distribution across the 150 faces of quantitative measurements for each of 32 quantitative characteristics.

Pairwise, correlations of average beauty scores between the pairs of secondary rater groups were assessed using Pearson's correlation coefficient. Within each rater group, for each facial characteristic the correlation between the average beauty score and quantitative measurement was assessed. The objective was to find the measurements that have the highest or lowest correlations with average beauty score.

The multivariate method, stepwise linear regression, was used to select the set of quantitative characteristics most predictive of average beauty score. Because of the high correlation between beauty scores for rater groups, the average scores from 100 raters were analyzed. Criteria for variable selection included assessment of the multiple correlation coefficient and application of a significance level of .05 for variable entry and retention. The objective was to choose the model with the highest multiple correlation coefficients with statistically significant coefficients for all predictors in the model.

Results

Figure 4, A–E are montages of the 30 faces created for each generation. Notably, in the parent generation, the faces are heterogeneous in distinct contrast to the later generations where there is profound convergence of features. The P and F1 generations demonstrate diversity with respect to most facial features. Faces are asymmetric and there is a wide variation in facial shape and proportion. With each successive generation, symmetry becomes more prevalent and clarity of skin increases, which is a product of image averaging. In the F3 and F4 generations (Fig. 4, D–E), oval faces clearly predominate; lips are fuller, and eyebrows more distinct and arched. There is significant similarity in terms of the size and shapes of the lips, nose, and eyes. All faces are symmetrical, and the brow shape

is arched and nearly identical to the “ideal” brow shape described in most plastic surgery and cosmetology texts. Notably, in the P (Fig. 4A) and F1 (Fig. 4B) images, some semblance of ethnic diversity is maintained, but the repetitive morphing process eliminates this with successive generations.

Figure 5 depicts the average attractiveness scores for each generation, P-F3 (white bars) and the average attractiveness scores of the subset of faces that were selected by the algorithm for morphing P_b -F3_b (darker, shaded bars), as determined by the original initial focus group of 17 trained student evaluators. In each successive generation and its corresponding breeding cohort, attractiveness scores increase each generation through the F3 generation. Notably, the standard deviation (SD) bars narrow slightly, thus further underscoring the convergence of features observed in Figure 4 D–E above. The initial focus group did not evaluate the F4 generation, as there was no intent to morph/breed, and the F5 generation.

Figure 6 depicts the average attractiveness scores for P-F4 (white bars) and the average attractiveness scores of the subset of faces that were selected by the algorithm for morphing P_b -F4_b (darker, shaded bars), as determined by the final focus group, which did evaluate the final generation (F4). The data represents the average attractiveness scores of all three final evaluation focus groups (undergraduates, surgeons, and cosmetology students) whose results were pooled for this analysis. Attractiveness scores increased with each generation. Notably, the average score for the P generation in Figure 6 was significantly lower than that of the initial student evaluator group illustrated in Figure 5. Histograms for each generation (Fig. 7 A–E) show the distribution of attractiveness scores and demonstrate the dramatic shift in terms of average score, but also show movement of the median and alteration in the skew. Each histogram shows the frequency of each score^{1–10} for a specific generation (total of approximately 3,000 votes per generation). The observation that the subset of faces which went on to morph or digitally “breed” in each case had a slightly higher beauty score for each generation, demonstrating the effect of introducing a selection pressure into the genetic algorithm. The most and least attractive face for each generation is depicted along with the corresponding score in Figure 8. Of note, with the later generations, the spread between attractive and unattractive faces narrows.

The morphometric measurements on the features identified in Figure 3 are listed in Table I. It must be emphasized that these are relative measurements in arbitrary units and are measured from images that have all been scaled so that the distance from the trichial line to the lowest point of the chin is the same in each image. Since it is generally acknowledged that there is at least a loose relationship between facial proportions or distances and attractiveness, statistical analysis was performed to determine whether any relationships existed between any of these measured values and attractiveness score. The final focus group evaluations were used for this analysis. Within rater groups, statistically significant correlations between average beauty score and quantitative measurements were identified only for three facial characteristics (nose width, right eyebrow peak, and upper left Cupid's bow), and these were notable for weak correlations (see Table II). Surprisingly, no correlations were identified for the more germane measurements such as nasal height, facial thirds, facial fifths etc. On the basis of 149 faces, pairwise Pearson correlation coefficients for average beauty score varied from 0.964 to 0.972 when pairs of the three rater groups were compared, strongly indicating that these groups define and evaluate beauty similarly despite different training and professional background.

Characteristics most predictive of average facial attractiveness score were selected using stepwise linear regression analysis. The model with three characteristics, the height of the upper left Cupid's bow ($P = .002$), the height of the right eyebrow arch ($P = .032$), and the height of the right eyebrow at its most medial point (0.031), was significant (overall model

F-value, 0.0003). The multiple correlation coefficient for this model was 0.12, indicating that 12% of the variability in average facial attractiveness score was explained by the regression on these predictors.

Discussion

In this study, realistic appearing synthetic facial images were created using morphing software across all generations. The key facial elements (eyes, lips, etc.) were distinctly preserved through each generation, though normal features of human skin such as blemishes, nevi, and acne were averaged out. Overall, with each successive generation, faces became more symmetric, and the overall appearance of the faces assumed a more multi-racial appearance with honey-colored skin, intermediate facial features, in contrast to previous reported studies, which usually focused on subjects of only European extraction. Likewise, vestiges of frank ethnicity drop out with the F2 generation. This investigation did not seek to examine or eliminate the impact of ethnicity in either the subject population that was photographed and used to produce morphs or in the focus groups used to score each synthetic image. Ongoing work in the lead researcher's group is currently focused on examining these factors. Parallel studies will examine only morphs derived from individuals of European heritage.

The montages (Fig. 4) demonstrate that the similarity in facial features increased with each successive generation. This suggests that the algorithm iterates to a solution or stable point, even when using the small sample size ($n = 30$) employed in this study. The observed general trends include a greater prevalence of oval shaped faces, fuller lips with distinct Cupid's bows, and defined and arched eyebrows. The nose does not form a prominent feature on any of the later generation faces. The improvement in overall attractiveness is supported by both the increase in average attractiveness score with each generation (Fig. 5 and Fig. 6) and the shift and change in median values and skew for the corresponding histograms (Fig. 7). The increase in the average value of more than one attractiveness point is a profound shift and indicative of the impact of using a genetic algorithm focused on cultivating beauty. The reduction in the SD of each average beauty score with each successive generation also underscores how this algorithm produces a modest degree of convergence as well.

On an individual basis, each F3 and F4 face is distinct and unique, but when examined collectively as in the montages, patterns and trends do emerge. The faces look eerily similar as they share virtually identical facial shape, lip fullness, nasal contour, and brow shape. This effect can, in part, be attributed to the innate averaging process that occurs with morphing¹⁹ in combination with the selection pressure exerted on the population using the genetic algorithm. Likewise, the use of a small parent population of only 30 faces can limit diversity and bias results. In such a small population, the impact of one or two extremely attractive faces may have a significant impact. For example, the F3 cohort was the product of three generations of algorithm execution. One very high-scoring F3 morph had the same "great-grandmother" on three separate branches of its family tree. In nature, classic examples of this effect are Darwin's finches where unique selection pressures, very small populations, isolation, and time led to very distinct species occupying unique niches.

The statistical analysis reveals some interesting quantitative trends. The attractiveness scores produced by final focus group (undergraduates, surgeons, and cosmetology students) correlated very well with one another, indicating general agreement in terms of what each group defines as facial attractiveness. Experts (surgeons and cosmetologists) rated the morphs the same as lay persons (undergraduates). The significant correlations between attractiveness score and the three facial measurements listed in Table II are intriguing. Nasal

width, the height of the right eyebrow arch, and the height of the left lower lip within the sagittal plane of the Cupid's bow peak all negatively correlated with facial attractiveness score. The identification of nasal width as a key factor agrees with what one would intuitively believe, namely that narrower noses are more attractive. The identification of laterality with respect to brow and lip dimensions is perplexing. In this setting, asymmetry of the brow and lip in a morphed face may be attractive over a face in symmetric repose because of some subtle cue related to a suggestive facial expression.²⁰ However, with the small sample size, this finding may be spurious and related to asymmetry in extremely attractive faces in the P generation, and the effect of this finding propagating with each successive generation. Surprisingly, facial attractiveness did not correlate with more traditional measures such as facial thirds or fifths.

The low magnitude of these correlations may be a consequence of several factors, including: 1) the small number of faces used in each generation; 2) the trend that images tended to converge in appearance with each successive generation; 3) the fact that only linear measurements were recorded; and 4) the fact that images were scaled to the same relative dimensions based on facial height, rather than focusing on absolute measurements. The first two points are important as with each successive generation, faces look more and more similar and attractiveness scores are high. There simply is less spread in the data than compared to a study with 150 randomly selected faces.

The morphing process at this time remains a labor-intensive endeavor requiring 30 to 60 minutes for each pair of images, and there is a substantial learning curve. Significant diligence is required around the eyes, eyelids, brows, and lips to achieve realistic morphs. Optimal morph construction requires attention to detail when constructing templates, and templates are best drawn using a very large monitor in combination with a digital graphics table, which affords finer control than either just a mouse or touchpad. The morphing process itself introduces artifacts in that average features result in increasing the clarity of skin, removing blemishes, altering color, and increasing symmetry. Also, in later generation morphs (i.e., F3 and F4), the features of the face are less sharp and distinct as if they are photographed using soft lighting and a diffusion filter. This effect alone may introduce a small bias in these later generations.

The similarity in appearance observed in the F3 and F4 generations may be a consequence of the algorithm iterating toward what might be considered the ideal face as determined by the initial focus group of evaluators. These 17 evaluators scored each of the faces, and these scores are the basis for the selection pressure within genetic algorithm. It is important to note that the scores of these 17 evaluators were used to generate the selection pressure, and that it was critical to use the same 17 for each successive generation. However, the small sample size of faces ($n = 30$) used in this study may introduce a bias to these results. Premature convergence to a non-optimal phenotype or, in mathematical parlance, convergence to a local maximum, may occur due to this sample size. Expanding the population to a larger number such as 300 subjects might aid in clarifying whether this indeed has occurred and is the focus of our ongoing investigations.

The genetic algorithm is not only limited by the size of the sample populations, but also by the biases intrinsic to the focus group used to assign attractiveness scores. In this study, the initial focus group determined "genetic" fitness. Ideally, focus group size should be massive to provide better reliability of scores. In this study, the initial focus group consisted of students enrolled in a seminar taught by the lead author. The disadvantage was: 1) the size of this group ($n = 17$); 2) the preponderance of women in the group; and 3) the fact that the ethnic composition of the evaluator group was not identical to the subject populations. The advantages were: 1) the same 17 evaluators saw each new generation of faces; 2) the

evaluators each had an individual visual analogue scale to aid in maintaining reproducibility of their scores; 3) to some degree, evaluators attempted to spread their scores across the scale in a logical rather than arbitrary manner; and 4) in theory, as the students were in a seminar focused on beauty and esthetic surgery, they have a more erudite approach to gauging attractiveness. By necessity, having the same group of evaluators is critical as 30 morphs take numerous man-hours to generate and a delay of 1 to 2 weeks was needed to morph each new generation. The second or final focus group that evaluated all 150 images at one setting were likely more prone to arbitrary scoring of faces and being randomly more or less charitable in attractiveness assessment (i.e., calling a modestly unattractive face a “1” and an attractive face a “10”). Currently we have ongoing investigations that compensate for focus group size and composition effects by using a novel Web-based approach to overcome these limitations.

Planned investigations will focus on increasing the sample population by a factor of 10, and also normalizing all facial dimensions with respect to the interpupillary distance. Regardless, there are numerous methods used to measure the face, many of which are more complex than the simple measurements used in this study. Despite the relative paucity of statistical data, in general, symmetry, oval-shaped faces, defined and arched brows, full lips, and small non-prominent noses remained consistent features in the highest rated faces, and at least on inspection, the rule of thirds and fifths is generally preserved.

The genetic algorithm in this study used facial attractiveness as the fitness function. The genetic representation is the appearance of each facial image. The selection process is a fitness-proportional selection model (also known as *roulette-wheel selection*) and is stochastic in that a small proportion of less attractive faces reproduce/morph in each round of algorithm execution. This approach enhances the diversity of the breeding populations, and presumably reduces the chance of premature convergence to a local maximum. Regardless, small sample sizes may still result in convergence to a local maxima rather than iterating toward a universal/global solution. There are obvious practical limitations in our approach in that only a microscopic subset of the U.S. female population is used, and time and manpower requirements reduce the number of morphs that can be created and the number of selection rounds in which to execute the algorithm. Perhaps one advantage of using a diverse multi-ethnic population in this study is that the facial features are quite diverse, enabling creation of “larger mutations” by morphing faces with very different features.

On the other hand, traditional approaches to identify “the perfect face” or defining facial beauty using quantitative methods have relied on correlating focus groups' facial attractiveness scores primarily with morphometric measurements. We propose that using our algorithm, a population of synthetic faces evolves and iterates toward at least a local maximum to provide a glimpse of the elusive perfect face.

Conclusions

The use of genetic algorithm in combination with morphing software and traditional focus group-derived attractiveness scores can be used to evolve attractive synthetic faces. We have demonstrated that the evolution of attractive faces can be mimicked in software. The approach creates a virtual “Galapagos,” with beauty acting as the selection pressure. Genetic algorithms and morphing differ substantially from traditional methods that rely heavily on correlating attractiveness scores with a series of morphometric measurements, and in the end, do not produce an ideal composite attractive face. Clearly, to fully exploit the potential advantages of this approach, research will require the development of automated software

algorithms to increase generation throughput, examination of lateral images, employment of larger basis populations, and examination of the impact of various demographic factors.

Acknowledgments

The authors would like to thank Adam Summers, PhD, for discussions on evolution and developing a genetic algorithm, Jon-Paul Pepper, MD, and Amir Karam, MD, for assistance with institutional review board issues, Preeya Desai for assistance with facial measurements, and Sylvia Allen of the COBA Academy.

Bibliography

1. Larrabee WF Jr. Facial beauty. Myth or reality? *Arch Otolaryngol Head Neck Surg.* 1997; 123:571–572. [PubMed: 9193214]
2. Pettijohn TF 2nd, Jungeberg BJ. Playboy playmate curves: changes in facial and body feature preferences across social and economic conditions. *Pers Soc Psychol Bull.* 2004; 30:1186–1197. [PubMed: 15359021]
3. Bashour M. History and current concepts in the analysis of facial attractiveness. *Plast Reconstr Surg.* 2006; 118:741–756. [PubMed: 16932186]
4. Edler R, Agarwal P, Wertheim D, Greenhill D. The use of anthropometric proportion indices in the measurement of facial attractiveness. *Eur J Orthod.* 2006; 28:274–281. [PubMed: 16415084]
5. Honn M, Goz G. The ideal of facial beauty: a review. *J Orofac Orthop.* 2007; 68:6–16. [PubMed: 17238049]
6. Farkas, LG.; Farkas, LG. *Anthropometry of the Head and Face.* New York: Raven Press; 1994. p. xix–405.
7. Etcoff, NL. *Survival of the Prettiest: The Science of Beauty.* New York: Doubleday; 1999. p. 325
8. Halberstadt J, Rhodes G. It's not just average faces that are attractive: computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychon Bull Rev.* 2003; 10:149–156. [PubMed: 12747502]
9. Rhodes G. The evolutionary psychology of facial beauty. *Annu Rev Psychol.* 2006; 57:199–226. [PubMed: 16318594]
10. Rhodes G, Yoshikawa S, Clark A, Lee K, McKay R, Akamatsu S. Attractiveness of facial averageness and symmetry in non-western cultures: in search of biologically based standards of beauty. *Perception.* 2001; 30:611–625. [PubMed: 11430245]
11. Cornwell RE, Law Smith MJ, Boothroyd LG, et al. Reproductive strategy, sexual development and attraction to facial characteristics. *Philos Trans R Soc Lond B Biol Sci.* 2006; 361:2143–2154. [PubMed: 17118929]
12. Smith MJ, Perrett DI, Jones BC, et al. Facial appearance is a cue to oestrogen levels in women. *Proc Biol Sci.* 2006; 273:135–140. [PubMed: 16555779]
13. Johnston, VS. *Why We Feel: The Science of Human Emotions.* Reading, MA: Perseus Books; 1999. p. ix–210.
14. Enquist M, Ghirlanda S. Evolutionary biology: the secrets of faces. *Nature.* 1998; 394:826–827. [PubMed: 9732862]
15. Langlois JH, Roggman LA. Attractive faces are only average. *Psychol Sci.* 1990; 1:115–121.
16. Johnston VS. Mate choice decisions: the role of facial beauty. *Trends Cogn Sci.* 2006; 10:9–13. [PubMed: 16311066]
17. Johnston VS, Solomon CJ, Gibson SJ, Pallares-Bejarano A. Human facial beauty: current theories and methodologies. *Arch Facial Plast Surg.* 2003; 5:371–377. [PubMed: 12975133]
18. Eienthal Y, Dror G, Ruppin E. Facial attractiveness: beauty and the machine. *Neural Comput.* 2006; 18:119–142. [PubMed: 16354383]
19. Perrett DI, May KA, Yoshikawa S. Facial shape and judgements of female attractiveness. *Nature.* 1994; 368:239–242. [PubMed: 8145822]
20. Swaddle JP, Cuthill IC. Asymmetry and human facial attractiveness: symmetry may not always be beautiful. *Proc Biol Sci.* 1995; 261:111–116. [PubMed: 7644543]

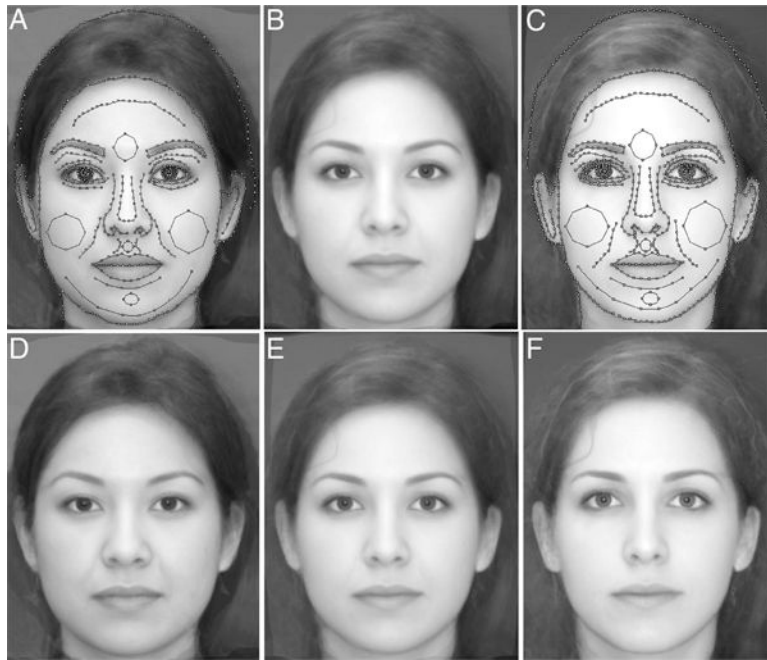


Fig. 1. Templates and registration points for generation morphs. Two facial images, A and C, are used to generate a 50:50 morph (B). Lines and regions of interest mark the key features that must be co-registered on both faces. D, E, and F show each face without overlying template.

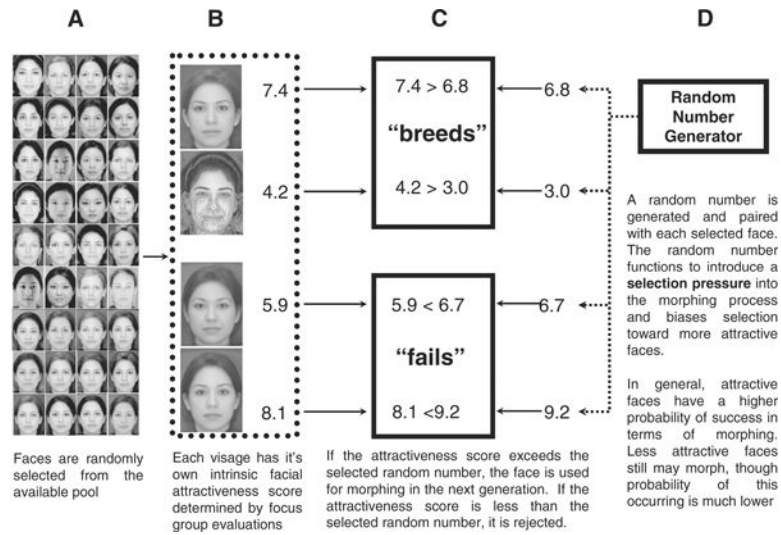


Fig. 2. Schematic of genetic algorithm for evolving attractive faces. Faces are randomly selected from a pool of available faces (A). Each visage has an intrinsic facial attractiveness score determined by a focus group (B). Attractiveness scores are compared (C) with numbers produced by a random number generation (D). Accordingly, attractive faces are successful (and then are used for morphing and creating the next generation of faces) if their attractiveness score exceeds the number generated by the random number generator. Likewise, faces fail if their attractiveness score is less than that produced by the random number generator.

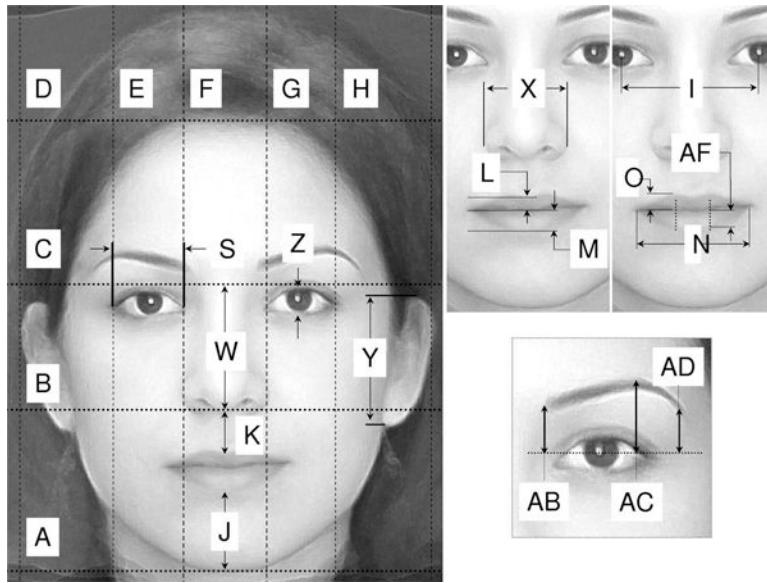


Fig. 3. Diagram of specific facial features measured on all evolved morphs. The letters on the diagram are described in detail in Table I. Note, not all symmetric measurements are labeled to preserve clarity.



Fig. 4.
A–E. Montages of morphed faces for parent, P (A), F1 (B), F2 (C), F3 (D) and F4 (E) generations.

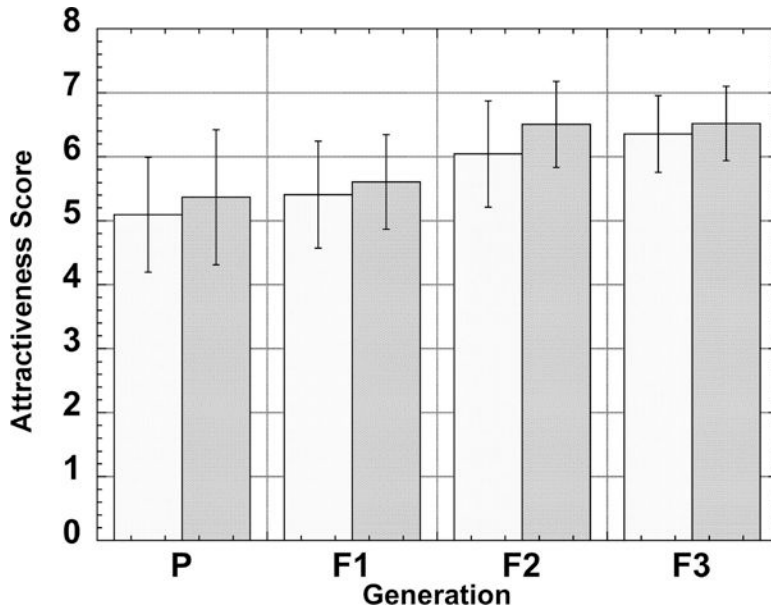


Fig. 5. Attractiveness score as a function of generation produced by initial focus group of evaluators. White bars indicate average scores in each generation. Gray bars represent the average scores of the subset of faces forming the breeding cohort for each generation. (The initial focus group did not score the fourth generation.)

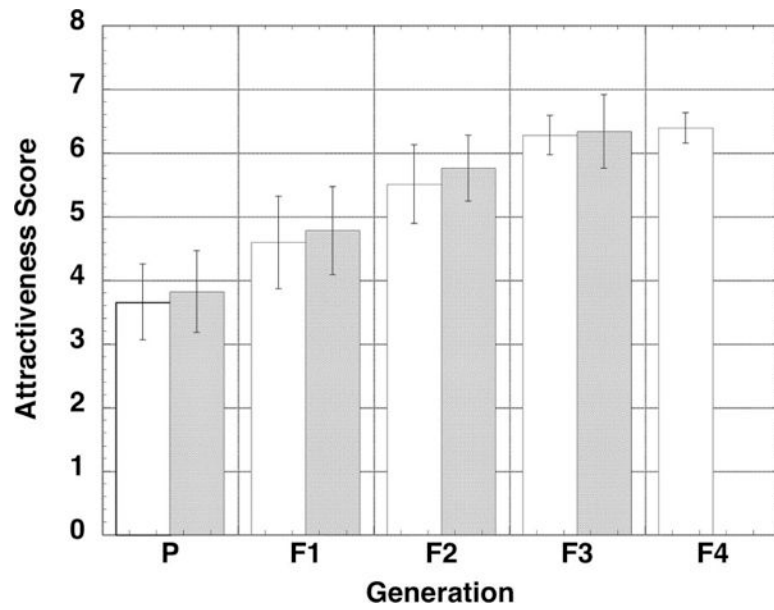


Fig. 6. Attractiveness score as a function of generation produced by final focus group of evaluators. White bars indicate average scores in each generation. Gray bars represent the average scores of the subset of faces forming the breeding cohort for each generation. No breeding cohort was selected from F4.

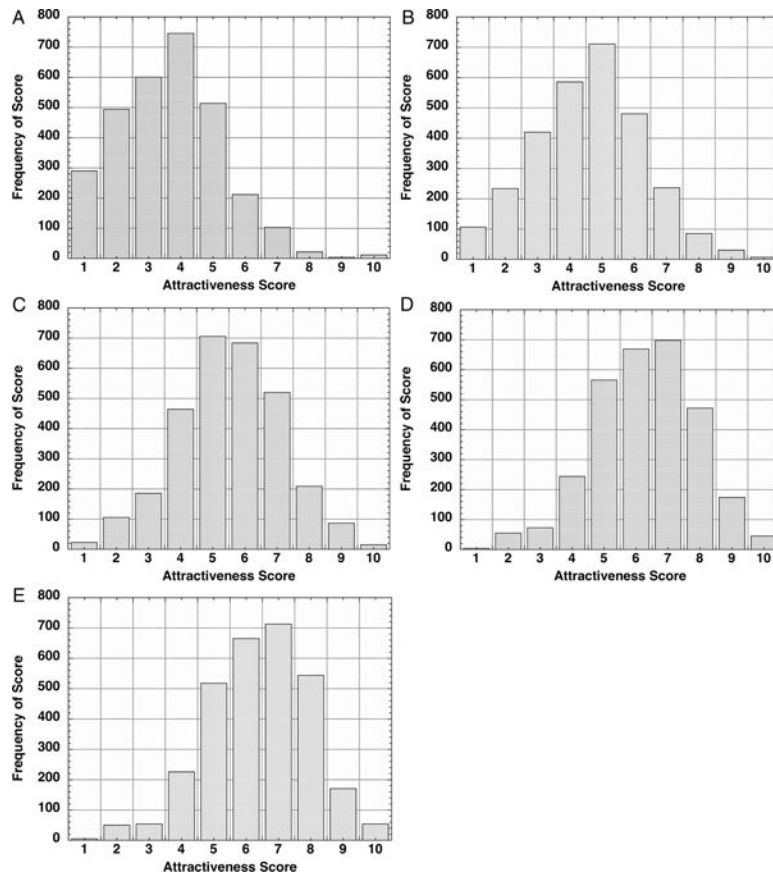


Fig. 7. A–E Histograms illustrating the distribution of attractiveness scores for parent (A), F1 (B), F2 (C), F3 (D), and F4 (E) generations. The total number of votes per generation was 3,000.

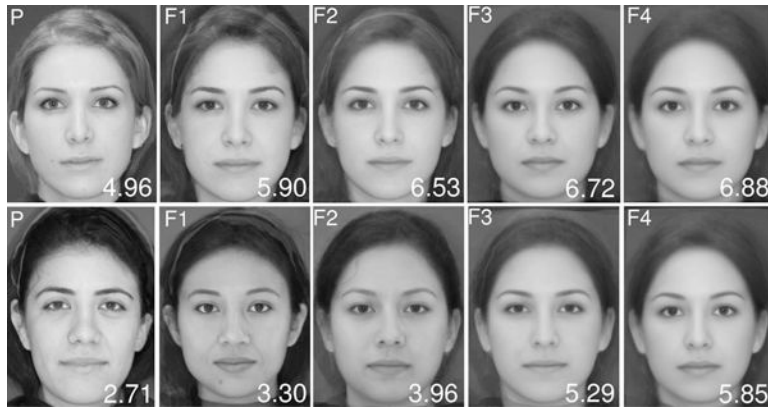


Fig. 8. Images with the highest (upper row) and lowest (bottom row) attractiveness score for each generation. The attractiveness score for each face is inset in the lower right corner.

Table 1
 Summary of Descriptive Statistics on Facial Measurements (Note: Measurements Are A.U.).

Key	Morphometric Measurements	Median	Mean	SD	Min.	Max.
A	Face height–lower third	77.6	76.6	4.262	61.2	86
B	Face height–middle third	63.55	62.6	3.91	48.8	69.7
C	Face height–upper third	79.72	79	3.615	65	85.9
D	Face width–left lateral	47.97	47.7	3.364	37.3	58.6
E	Face width–left intermediate	31.91	31.8	1.887	25	42
F	Face width–central	38.18	37.8	2.54	17.7	43.2
G	Face width–right intermediate	32.21	32.1	2.231	25.4	43.1
H	Face width–right lateral	47.73	47.6	3.549	36.8	60.2
I	Interpupillary distance	71.47	71	2.805	62	76.6
J	Lip height–lower lip to menton	39.05	39	3.36	19.8	58.2
K	Lip height–subnasale to central vermillion border	19.49	19.4	1.396	16.3	24.4
L	Lip height–upper lip–vermillion to stomion	6.47	6.45	0.933	3.48	9.66
M	Lip height–stomion to vermillion	12.05	12.2	1.071	9.54	15.1
N	Lip width (inter-commissure)	60.14	59.9	3.86	40.8	67.7
O	Lip height–lower left Cupid's bow (at the vertical plane of Cupid's bow peak)	11.64	11.7	1.055	9.51	15.2
P	Lip height–lower right Cupid's bow (at the vertical plane of Cupid's bow peak)	11.26	11.5	1.933	8.42	31
Q	Left ear height	62.68	62.2	4.77	48.7	79.3
R	Left eye height	13.11	13	1.109	10	15.5
S	Left eye width	31.9	31.8	1.879	25	42
T	Left eyebrow height relative to mid-pupillary line–most medial point	15.34	15.2	1.484	9.8	17.9
U	Left eyebrow height relative to mid-pupillary line–at brow arch	13.99	13.9	1.985	8.79	21.7
V	Left eyebrow height relative to mid-pupillary line–most lateral point	20.89	20.8	2.278	10.6	30.3
W	Nose height–nasion to subnasale	63.55	62.6	3.91	48.8	69.7
X	Nose width–between alar creases	41.71	42.3	3.631	32.7	69.5
Y	Right ear height	60.96	60.3	5.443	29.3	73
Z	Right eye height	12.63	12.7	1.114	10.4	15.3
AA	Right eye width	32.21	32	2.049	25.4	40
AB	Right eyebrow height relative to mid-pupillary line–most medial point	15.32	15.3	1.584	11	18.9
AC	Right eyebrow height relative to mid-pupillary line–at brow arch	14.14	14.2	2.402	7.41	22.1

Key	Morphometric Measurements	Median	Mean	SD	Min.	Max.
AD	Right eyebrow height relative to mid-pupillary line–most lateral point	21.52	21.6	1.92	17.2	26.9
AE	Lip height–upper left Cupid's bow-peak of bow to stomion	8.31	8.28	1.042	4.23	10.7
AF	Lip height–upper right Cupid's bow-peak of bow to stomion	8.33	8.29	0.928	5.04	11.2

SD = standard deviation; Min. = minimum; Max. = maximum.

Table II

Summary of Pearson Correlation Coefficients Between Facial Attractiveness Scores and Quantitative Measurements.

Variable	Average Beauty Scores		
	Cosmetologists	Surgeons	Undergraduates
Nose width between alar creases (X)	-0.21 *	-0.208 *	-0.213
Right eyebrow height relative to mid-pupillary line-at brow arch (AC)	-0.187 *	-0.226	-0.197 *
Lip height-upper left Cupid's bow-peak of bow to stomion (AE)	-0.243	-0.254	-0.238

.0001 *P* value < .01;

* .01 *P* value < .05.