When beauty breaks down: Investigation of the effect of spatial quantisation on aesthetic evaluation of facial images

Talis Bachmann
Department of Psychology and Institute of Law, University of Tartu, Kaarli puiestee 3, Tallinn 10119, Estonia; e-mail: talis.b@lawinst.ee
Received 23 August 2005, in revised form 13 October 2006; published online 8 June 2007

Abstract. Research on the perception of facial attractiveness has been dominated by aspects of averageness, symmetry, and secondary sex characteristics of faces. Almost absent is systematic research on the spatial scale (coarseness) of detail sufficient to carry information about facial attractiveness. In the present study, subjects were asked to evaluate the attractiveness of faces while the coarseness of detail of the face images was systematically decreased. Subjects discriminated a set of attractive faces from the set of unattractive faces when the coarseness of spatial quantisation changed from 10 to 17 pixels per face. Some of the faces regarded as attractive had the steepest shift in the rating towards the more attractive end of the scale at different resolutions. Once perceived as attractive in the coarse scale of image resolution, the faces generally did not return to the unattractive group with a subsequent finer scale of description. No single critical scale of detail was revealed that would have dramatically changed the perception of facial attractiveness from uncertain to distinct.

1 Introduction
To say that a face is beautiful (or not) requires that the information carried by a facial image is sufficiently rich and this applies to beautiful and ugly physiognomies which are both psychophysically ‘rich’ in their own respects. Research on facial beauty, or ‘attractiveness’ as it is usually termed, has focused primarily on a few important factors of phenotypal morphological characteristics such as symmetry, averageness, racial attributes, health, and secondary sex characteristics including hormone markers (eg Thornhill and Gangestad 1993, 1999; Perrett et al 1994, 1998; Fink and Penton-Voak 2002; Little and Perrett 2002; Rhodes and Zebrowitz 2002; Rhodes et al 2003, 2005a, 2005b). In this context, much less is known about the role of some basic psychophysical measures—the principal carriers of aesthetic appeal, which have been thoroughly studied in mainstream psychophysics. One such prevailing dimension dominant in psychophysics, but overlooked in facial-attractiveness research, is the scale of visual detail.

Typically, when spatial-scale factors that are important in face identification/recognition are studied, spatial detail is measured and manipulated by variable spatial frequency or coarseness of spatial quantisation (eg Harmon and Julesz 1973; Watt 1988; Bhatia et al 1995; Hughes et al 1996; Lander et al 2001; Morrison and Schyns 2001; Sinha 2002; Bruce et al 2003). Even so, in spite of research on what kind of information carries identity, questions remain unanswered such as whether or not attractiveness is carried by coarse-scale or fine-scale visual information. Is there a particular narrow range of image detail that must be included in facial images so that their attractiveness can be evaluated, or do different faces carry attractiveness features at different spatial scales? Can a face be perceived as beautiful at one level of spatial detail, but not at another level? Applied questions are also important. In practicing and developing animation, advertising, computer games, optional illustrations of texts, video conferencing, videophones, etc one may want to know when attractiveness can no longer be displayed because image resolution has lost the necessary level of detail. Or, following just the opposite aim, how not-so-attractive original images could be depicted
as ‘bearable’ in terms of attractiveness, but without compromising recognisability when coarsening the image?

The present exploratory study aimed to examine the effects of spatial quantisation of facial images on psychophysical ratings of their aesthetic (attractiveness-related) value. (Following the prevailing tradition, I use the concept of attractiveness interchangeably with another parallel concept—beauty. But see, e.g., Donovan 2003 about the pitfalls in confusing ‘beauty’ with ‘attraction’. Indeed, a not so beautiful person can have virtually magnetic attraction for some others, especially in the political and sexual contexts of human interaction.)

If we look at a pair of coarse-quantised images of female faces (figure 1), it is not obvious whether they are derived from original faces for which there is a general consensus that they are attractive. (Actually, these faces belong to the polar ends of the attractiveness-ranking scale, provided that their fine-quantised versions and not coarse-scale versions were rated by experimental subjects.) But how coarse does the image have to be for perceived attractiveness to break down?

In order to find out the value of the scale of spatial degradation of facial images that would affect facial attractiveness, the value of pixelation of the images of faces was systematically varied while the subjects rated the perceived attractiveness of the faces. A set of different faces was used in order to establish whether different faces acquire their aesthetic determinateness at the same psychophysical coarseness value established after a critical step of de-coarsening the image, or whether they undergo a change in attractiveness at different individual critical values of coarseness.

**Figure 1.** Examples of two coarse-quantised face images that are difficult to evaluate according to how attractive (beautiful) the respective original faces could be. The left one belongs to face N that was rated as very attractive (mean rating 8.5) when displayed as a fine-quantised image. The right one belongs to face K that was rated as unattractive (mean rating 4.7) when displayed as a fine-quantised image. A colour version of this figure is shown on the Perception website at [http://www.perceptionweb.com/misc/p5509/](http://www.perceptionweb.com/misc/p5509/).
2 Experiment

The method used involved spatial quantisation of the original set of facial images. 14 female faces were randomly selected from the set of faces collected from various public-access sources on the internet. These included average, attractive, unattractive, and randomly sampled face images. This selection can be regarded as quasi-random with regard to the public-domain internet sources. Each one of these faces was spatially quantised (pixelised) by IrfanView freeware at seven different levels of quantisation. In the main experiment, the quantised faces were shown on a white screen, for 1.5 s each, to a group of thirty-six subjects who rated their apparent attractiveness on a 10-point rating scale. In an independent supplementary experiment, and in order to be able to determine the distribution of the sampled faces between different levels of attractiveness ratings in the conditions where original images are used, the non-quantised versions of all faces were shown to a different group of thirty-six students (twenty-five females, gender distribution proportionally and approximately representative of the student population of the school) who rated how attractive the faces appeared to them on a 10-point rating scale. As a result of this, two groups of faces were formed: those with mean ratings above the median 5.5 points of the scale (a set of 5 faces perceived as more attractive than median: B, C, L, M, N), and those with mean ratings below 5.5 points (a set of 9 faces perceived as less attractive than median: A, D, E, F, G, H, I, J, K)—see figure 2.

A comparison of female and male sub-population data showed that there were no statistically significant differences between mean ratings produced (females: mean = 5.14, SD = 2.15, SEM = 0.57; males: mean = 4.75, SD = 1.97, SEM = 0.53; t_{26} = 0.508, p = 0.62). The Pearson coefficient (r = 0.917, p = 0.001) showed highly significant correlation between female and male mean evaluations of the same faces. Further analysis of gender differences was therefore not pursued.

**Figure 2.** Mean ratings of facial attractiveness of the original faces by the group of independent evaluators in the supplementary experiment. The median level of ratings on the 10-point scale splits faces conditionally into two groups: attractive faces (N, L, M, B, C with ratings 8.5, 8.2, 7.4, 6.8, and 6.0, respectively) and not so attractive faces (I, K, H, J, E, F, G, A, D with ratings 5.4, 4.7, 4.3, 3.9, 3.7, 3.4, 3.0, 2.8, and 2.1, respectively).
2.1 Methods

2.1.1 Participants. Thirty-six volunteer subjects aged 19–27 years took part in the main experiment; sixteen were male and twenty were female. All were students of law and psychology who participated in partial fulfillment of course credits. All subjects had normal or corrected-to-normal vision. All were naive as to the purpose of the experiment when it was conducted, but they were debriefed and participated in subsequent discussions.

2.1.2 Stimuli and apparatus. 14 original frontal-face images of the quasi-randomly chosen female picture-portraits, labelled from A to N were first resized to 200 × 305 pixels rectangular area and transformed into gray-scale images. If necessary, they were cropped, rescaled, and framed so that a set of frontal full-face images of equal size could be used. The top of the head and bottom of the chin marked the uppermost and lowermost points of the image, respectively. The finest level of depiction was abandoned for the stimuli used in the main experiment and the finest spatial resolution allowed in the experimental images was 5 screen-pixels per one isoluminant square-shaped block (effective pixel) of the experimental test image. Because the picture quality of the original faces was not equal, carrying out the blocking-pixelation which was more coarse than pixelation of the original images made it possible to smooth out and equalise differences in the fine-grain quality of local detail and remove the minute fine-grain skin defects in all experimental face images, without losing the generally acceptable face-image quality at the finest level of quantisation (40 pixels per image).

Each image was spatially quantised by the pixelising operation at seven different values of quantisation: 10 blocks per image along the horizontal interauricular dimension (the coarsest quantised image), and 12, 14, 17, 20, 25, and 40 blocks per image (the finest quantised image). Thus, overall, \(7 \times 14 = 98\) test images were prepared. They were arranged into a quasi-random sequence with decreasing scale of coarseness in a Powerpoint (ppt) slides presentation file for use in the experimental presentations.

The quantised face stimuli were presented at varying viewing distances for different subjects, the distances ranging from 4 to 7.5 m. The vertical dimension of the facial test images presented on a white screen in a moderately lit lecture room (with ascending height of the rows of the seats) was 55 cm. The images of faces were presented at the centre of the screen. (From the later analysis of the experimental results it appeared that the variable size of the stimulus images for different subjects did not have any substantial qualitative effects, probably owing to the fact that the experimental variable of scale of quantisation had its effect in terms of image coarseness relative to the size of the face image. Because this measure was invariant for all different subjects’ viewpoints, the present method of group experiment is acceptable.)

2.1.3 Procedure. The experiment was run simultaneously on a group of thirty-six subjects (twenty females). First, it was explained to the whole group that a set of spatially quantised pictures of female faces will be shown, each for 1.5 s, and 10 s will pass between the presentations of successive pictures. [Sufficiently clear perception of spatially degraded images requires sufficient exposure duration so that all levels of detail can be perceptually integrated (see, eg, Parker et al 1997). Exposure durations that are too short may lead to uncontrolled artifacts. The duration of 1.5 s is sufficient for all spatial levels of detail to be integrated and therefore possible artifacts of short duration are avoided; see Bachmann (1991) about the time-course of perception of spatially quantised facial images] An example of a quantised face was also shown to the subject group. The subjects were instructed that a rating of attractiveness of each stimulus face had to be produced on a prepared response sheet. They were asked to rate all stimuli; no empty cells on the response sheets were allowed. A 10-point rating scale was used, with the least attractive face in subject’s opinion rated 1 and the most
attractive face rated 10. (Any assumptions about the nature of the possible 'function of attractiveness' and about its generality are deliberately avoided here.) Thus, a direct-scaling-like approach was used with ‘attractiveness’ being the attribute for which magnitude had to be estimated. An additional deviation from the more rigorous Stevensonian approach was that standards were imaginary (not real stimuli). This method was used in order to avoid the possibility that a face used as a stimulus in the experiment could be estimated as being more attractive than a real standard of attractiveness if it had been used.

After reading the instructions, subjects were allowed to ask all necessary questions about the task and procedure and the experimental requirements they had to follow. Stimuli were presented in the order of decreasing coarseness of quantisation (increasing fine-scale depiction), beginning with 10 blocks per face presentation (block size 20 screen-pixels) and ending with 40 blocks per face presentation (block size 5 screen-pixels). However, within each level of coarseness, the order of presentation of different faces was random. In this way, in addition, the faces in both above- and below-median attractiveness conditions were presented in random order. Before each face stimulus exposure, a warning announcement (“attention”) was given, which came 1.5 s before the face. The onset asynchrony between the stimulus faces was 10 s during which subjects wrote down their ratings of how attractive they judged the previously presented quantised face. At some point, during later presentations when quantisation scale reached higher detail, subjects began to understand that some differently quantised faces were actually derived from the same original face. However, the post-experiment interviews with subjects and the generally universal change of attractiveness ratings even on moving from the last-to-last to the very last (finest) scale level of quantisation ensured that subjects were able to ignore the potential effects of face repetition and were able to concentrate on perceptual aesthetic impression left by the face. (Moreover, no precisely identical stimulus faces were repeated because there was a slightly varying appearance of the same original face. This was due to the difference in the configuration in the mosaic of quantisation blocks even between the relatively fine neighbouring scales of quantisation. With coarse and intermediate scales of quantisation in the initial and middle parts of the experiment there were no subjects who could have noticed the repetition of faces.)

3 Results
Total mean of ratings for stimuli with variable levels of quantisation ranged from 3.71 (SD = 2.23) for 10-blocks-per-face stimuli to 4.84 (SD = 2.45) for 20-blocks-per-face stimuli. Getting higher attractiveness ratings for more detailed (less coarse) faces, as seen on figure 3, is due to the ratings of a set of faces that were perceived as more attractive than the median by an independent group in the supplementary experiment (faces B, C, L, M, N rated above 5.5 points on the average). The group of faces rated independently as unattractive (A, D, E, F, G, H, I, J, K) did not contribute to this increase: they were perceived as fairly constantly unattractive with increased pixel number. The subjects apparently anchored their rating scale at the beginning when little information about the attractiveness quality was available (accessible?) at the unattractive part of the scale.

An analysis of variance showed that both the effect of attractiveness group and the effect of level of quantisation were highly significant ($F_{1,12} = 46.21$, $p < 0.0001$; $F_{6,12} = 14.06$, $p < 0.0001$, respectively). Even more important is acknowledgment of the highly significant interaction between the level of quantisation and the type of stimulus ($F_{6,12} = 14.25$, $p < 0.0001$). This interaction and the graphs depicting mean ratings of faces belonging to two attractiveness groups as a function of coarseness of quantisation (see figure 3) emphasise the main findings of this study. Varying coarseness causes
a gradual increase in the attractiveness of highly attractive faces but not in that of less attractive faces. This is supported by a highly significant linear interaction between coarseness and attractiveness ($F_{1, 12} = 30.44, p < 0.0001$).

There is no particular level of quantisation that would decisively change the capacity of an image to carry information about its beauty—the change is gradual. When individual faces were examined one by one, the data showed that at quantisation levels of 10 and 12 pixels per face only one facial image (L) was rated as more attractive than the rest; the latter were perceived as less than average in terms of attractiveness (95% confidence limit). At a quantisation level of 17 pixels per face four faces (B, L, M, N) were perceived as more than average in attractiveness and the rest were less than average. At 20 and 25 pixels per face quantisation the same sets stood apart. Face C which was perceived as less than average up to 25 pixels per face level finally and abruptly moved to the attractive face category at the quantisation level of 40 pixels per face.

There was an indication of a possible categorical-perception effect occurring with an increase in the detail of depiction of facial images, which took place between attractive and unattractive faces. At the finest quantisation level, the smallest value of the rating of a face that was the least attractive one in the group of the attractive faces (face C: mean = 6.42, SD = 2.27, lower 95% confidence limit at 6.05) greatly exceeded the highest value of the rating of a face that was the relatively most attractive one in the group of the unattractive faces (face I: mean = 4.11, SD = 1.48, upper 95% confidence limit at 4.4). The difference was more than 2 scale points or 20% of the whole range of the rating scale and different from the corresponding comparison in the results of the independent group of raters in the supplementary experiment.

4 Discussion
The purpose of this study was to gather initial data concerning spatial-scale effects on the perception of facial attractiveness (beauty). It also sought to ascertain whether the capacity to discriminate attractive and unattractive faces depends critically upon
some narrow range of spatial-scale values. With coarse-quantised face images at 10 pixels per face, the perceptual value of attractiveness was degraded. Subjects could not clearly discriminate attractive and unattractive faces (with the exception of one face, L). At the finest scale of quantisation, with 40 blocks per face, subjects conspicuously differentiate attractive faces from unattractive faces (the difference between the upper and lower extreme values of means of ratings of the individual faces was almost 7 scale points: 8.69 – 1.94 = 6.75). The results showed a variation in the mean rating differences which were negligible at first, but diverged, gradually becoming highly significant with systematic increase in the detail of facial images. This interaction is due to the increase in the perceived attractiveness of the group of attractive faces while the faces belonging to the unattractive group remained perceived as flatly unattractive.

The more-or-less reliable bifurcation of the graphs representing the mean ratings of faces belonging to two subsets pertaining, respectively, to the group of attractive perceived faces and to the group of unattractive perceived faces takes place between quantisation values of 12 blocks (pixels) per face and 17 blocks per face. Interestingly, the value of this critical step in quantisation values is similar to the step that usually leads to an abrupt increase in the rate of correct identification of faces (Bachmann 1991; Bhatia et al 1995; Bachmann and Kahusk 1997). This seems to suggest that facial beauty is strongly associated with the facial characteristics that enable establishment of facial identity. If identity is difficult to extract, aesthetic quality—coincidentally—is also indeterminate. [However, see Rhodes et al (2003) about adaptation-dependent recalibration of preferential face prototypes, which allows for some share of independence of attractiveness perception from particular physical characteristics of an actual face.]

In the microgenetic course of the perception of facial attractiveness following repetitive presentations with systematically increasing resolution, the first moment of apprehension that some images are more attractive than others is arrived at a surprising coarse level of depiction—after 12 pixels per face width. Once this bifurcation of the two sets of functions has begun (see figure 3), almost no faces ‘fell back’ to the less attractive category when further increases in image detail were introduced. Because of this, it should be concluded that the physical carriers of facial beauty are originally present in the coarse configural properties of faces. [Analogously, in the microgenetic process of how artistic paintings are perceived the subjects can discriminate perceptual qualities pertaining to different artistic styles beginning from the very early microgenetic stages—Bachmann and Vipper (1983). Beauty can be present in the roots of perceptual microgenesis, based on coarse-scale gist of the image.]

There was generally no exchange between the positions of a face in the perceived ‘attractiveness space’ in the sense that the same face would belong to the attractive group at one level of quantisation and to the unattractive group at some other level of quantisation. Once perceived as attractive (or not so), a particular face remained in that category with further changes in the scale of coarseness. There were a few exceptions, however: face C ‘jumped’ from the unattractive rated group to the attractive rated group on quantisation change from 25 blocks per face to 40 blocks per face; face B underwent an analogous ‘jump’ between the 14 blocks per face and 17 blocks per face conditions. This means (and, particularly, taking into account the relatively small size of our sample of faces) that there is no universal precise value of spatial quantisation for all facial images where attractiveness or beauty ‘suddenly arrives’ or abruptly ‘breaks down’. Why some images acquire subjective quality of distinctly higher attractiveness at finer scale of depiction and others already at the coarse scale of depiction is a promising subject for research in the future.

An image with a quite poor quality of depiction (eg an image quantised at 40 blocks per face or coarser) can be reliably identified as attractive or not. The subtleties of eyes, mouth, nostrils, and other elements of a facial image cannot be perfectly discriminated
with this level of coarseness. When we cut out an eye or mouth region from such an image and show it separately to observers, they cannot evaluate the beauty of such poorly depicted images. Therefore, facial attractiveness can be encoded by some more general metrics used in describing configural properties of a face. Is it a measure of the balance of the distribution of elements, of a correspondence to culturally imprinted beauty standards, or of correspondence to the configural gene markers of good health?

An intriguing result revealed at the end of the rating session was that with fine-scale depiction of the images, an effect reminiscent of categorical perception emerged. Whether this is a true categorical-perception outcome or something else cannot be evaluated, given our data and methods.\(^{(1)}\),\(^{(2)}\)

Several factors have been tested in attractiveness research in earlier studies, which, because of the differences in the methods used, cannot be thoroughly discussed here: some of the variables would not have had any substantial causal role in producing our present results and the effects of some other variables were not subjected to experimental control. The effect of frontoparallelness of the faces, precisely controlled symmetry, height of the internal features of faces, relative luminance of different facial features, the quality and characteristics of skin, cannot be addressed here. [For this purpose, see, eg, Evans et al (2000); Sasaki et al (2005); Thornhill and Gangestad (1999); Little and Perrett (2002); Tyler (1998); Geldart et al (1999a); Russell (2003); Jones et al (2004).] The possible involvement of symmetry of face images and/or the possibility to keep some critical facial element unambiguously in the centre, as well as the possible effects of relative luminance of different facial features, all this is discussed in the footnotes.\(^{(3)}\),\(^{(4)}\)

The factor of the quality and characteristics of skin should remain out of consideration insofar as the local fine-grain texture aspect is of interest. The option to

\(^{(1)}\) This result is intriguing. The within-group differences between faces were much less on the average and in the evaluations of the original nonquantised images any hint at a possible categorical-perception effect was absent. Whether we deal here with categorical perception of facial beauty gradually developing in the course of ratings produced for the faces with every increasing detail of depiction, or with a sampling error (or variability) between the groups of subjects remains an open question. Further studies should take more care in sampling the faces and subjects. High-proof randomness of selection from a general population should be guaranteed and possible sampling errors caused by the deliberate motivations of posting the faces on internet and a small sample of raters should be avoided. Quite probably, internet sources may include relatively more attractive and less-than-average attractive faces, which may cause some bimodality in the respective distribution in internet. This may have flawed the present results. On the other hand, there is an intriguing different sociobiological theoretical possibility. Actually, the attractiveness-related phenotypal morphological characteristics may themselves be bimodally distributed.

\(^{(2)}\) In a related study, Schiano et al (2004) found that affective characteristics of faces appear not to be categorically perceived. But this should not mean that aesthetic qualities must follow the same rule.

\(^{(3)}\) The quantisation procedure used here was applied not as centred on images but as beginning the pixelisation-count from one side of the image for all different values of quantisation. This means that symmetry did not increase uniformly and monotonically with decrease in the coarseness of quantisation (although the aesthetic appeal of the attractive faces generally did). Moreover, the original faces were not carefully chosen according to whether they were symmetrically centred and their rotation with regard to the frontal plane perfectly adjusted or normalised. Some findings show that facial symmetry may not be the overwhelming characteristic of which faces are preferred and which are not because attractiveness can be evaluated from profile and half-faces (Scheib et al 1999; Valentine et al 2004).

\(^{(4)}\) Hypothetically, a critical value of image quantisation could exist where the size of pixel-blocks permits more-or-less unambiguous assignment of the gradients of the raw contrast of images to one or another part (feature) of the face. Bifurcation of the two groups of ratings—for attractive and unattractive faces—with changing quantisation could have taken place just when the assignment of luminance contrasts to the key facial features became meaningfully and harmoniously possible (see Bachmann and Kahusk 1997). An analogous commentary may be used also for the factor of perceived eye size (eg Geldart et al 1996b).
limit image spatial resolution at the upper end of the quantisation scales to 5 screen-pixels per block and not less (and also given the not too large size of the whole images) made the fine-grain skin-texture factor unimportant.

In conclusion, this exploratory experiment showed a couple of notable things. First, facial attractiveness can be reliably extracted from quite coarse degraded images (roughly about 15 blocks per face or an equivalent of 7 cycles deg⁻¹). Once established, face perception will not reverse into ‘unattractive’ with finer scale of depiction. Second, there is no universal precise value or universally narrow range of the scale of quantisation where all faces immediately lose or acquire their perceptual beauty. Third, it should be interesting to explore whether the bimodality or categorical-perception effect for attractiveness which was hinted by the present results may be either an artifact of sampling, a result of some prevailing rating strategy, or an expression of some more basic way of the bimodal distribution of the biological morphological characteristics.

Attractiveness is not only a pleasant thing, it also matters sociobiologically. It is useful to be able to detect its presence or absence at once. Therefore, a coarse level of human depiction is sufficient for the discrimination of attractiveness.

Acknowledgments. This study was in part supported by Estonian Science Foundation (grant #5778). Help from Iiris Luiga and Mark Elliott during preparation of this paper is gratefully acknowledged.

References
Bachmann T, Kahusk N, 1997 “The effects of coarseness of quantisation, exposure duration, and selective spatial attention on the perception of spatially quantised (‘blocked’) visual images” Perception 26 1181 – 1196
Bachmann T, Vipper K, 1983 “Perceptual rating of paintings from different artistic styles as a function of semantic differential scales and exposure time” Archiv für Psychologie 135 149 – 161
Geldart S, Maurer D, Henderson H, 1999a “Effects of the height of the internal features of faces on adults’ aesthetic ratings and 5-month-olds’ looking times” Perception 28 839 – 850
Geldart S, Maurer D, Carney K, 1999b “Effects of eye size on adults’ aesthetic ratings of faces and 5-month-olds’ looking times” Perception 28 361 – 374
Jones B C, Little A C, Burt M, Perrett D I, 2004 “When facial attractiveness is only skin deep” Perception 33 569 – 576
Little A, Perrett D, 2002 “Putting beauty back in the eye of the beholder” The Psychologist 15 28 – 32
Parker D M, Lishman J R, Hughes J, 1997 “Evidence for the view that temporospatial integration in vision is temporally anisotropic” Perception 26 1169 – 1180
Tyler C W, 1998 “Painters centre one eye in portraits” Nature 392 877 – 878
Conditions of use. This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.