Facial and vocal attractiveness:
a developmental and cross-modality study

Tamsin Saxton

Supervised by Dr P.G. Caryl

A dissertation submitted in fulfilment of the requirements
for the degree of

Master of Science in the Evolution of Language and Cognition

to

The Department of Theoretical and Applied Linguistics
The University of Edinburgh

2005

Acknowledgments

With grateful thanks to my supervisor Dr P.G. Caryl for his comprehensive and diligent guidance.

Abstract

Research on physical attraction is often interpreted with reference to the theory that physical attraction enables adaptive mate choice behaviour. The sensation of physical attraction is thought to enable humans to select those mates who are most likely to help them bear the fittest offspring. The sexual behaviour associated with mate choice emerges at puberty, and so the present study investigated whether adult-like judgments of facial and vocal attractiveness arise at puberty. It found that children and adolescents differ from adults in their judgments of attractive faces and voices, and that pitch of voice cues different responses in the different age groups. It also found co-variance in the attractiveness of male faces and voices, suggesting that the modalities of face and voice are providing concordant signals as to mate quality.

Contents

List of Tables and Figures

1. Introduction and Review

1.1 Introduction

The topic of physical attractiveness within humans has motivated wide-ranging empirical and theoretic research. Studies in this area are often interpreted with reference to the theory that physical attraction enables mate choice. The sexual behaviour associated with mate choice emerges at puberty, and so the present study asks whether adult-like judgments of attractiveness also arise at puberty: that is, do we find differences when comparing judgments about the attractiveness of different face and voices made by adults with those made by adolescents and children? It also attempts to replicate and extend work on the co-variance of facial and vocal attractiveness, and questions the extent to which one proposed predictor of vocal attractiveness – that of a low-pitched voice – affects judgments of vocal attractiveness in the different age groups. The experiment which was carried out is described at 1.4 below, following a review of the relevant literature on human attraction and mate choice.

1.2 Physical attractiveness

One of the most notable findings from work on human mate choice is that there are universal standards on what constitutes attractiveness. Extensive research has shown that, in general terms, people of all ages, sexes, sexual orientations and cultures agree on what constitutes an attractive face (overviews in Little, & Perrett, 2002; Langlois *et al* 2000; Grammer, Fink, Møller & Thornhill, 2003; Cunningham, Roberts, Barbee, Druen, & Wu,1995); someone of, say, Asian ethnicity will judge the relative attractiveness of two faces of European derivation in much the same way as a European judge would, and vice versa. Likewise, there is fairly strong agreement on what constitutes an attractive voice (eg Collins, 2000; Hughes, Harrison, & Gallup, 2002; Zuckerman, & Driver, 1989; Zuckerman, Hodgins, & Miyake,1990) and an attractive body scent (Thornhill, & Gangestad, 1999a; Rikowski, & Grammer, 1999; Gangestad, & Thornhill, 1998; Gangestad, & Thornhill, 1999), although cross-cultural studies are yet to appear.

Clearly, many aspects of physical attractiveness must be learnt or culturally acquired (for a review, see Zebrowitz, & Rhodes, 2000, cited in Rhodes *et al* 2001). For example, culture has an important influence on perceptions of ideal size, and members of different societies differ significantly in the body mass index (relating to the amount of body fat) which they choose as most attractive (Anderson, Crawford, Nadean, & Lindberg, 1992). At the level of individual variation, Cornwall *et al* (2004) summarise findings showing that female preference for heavily masculinized faces is influenced by the status of their relationship (Little, & Perrett, 2002), the age of their parents (Perrett *et al* 2002) and their self-rated attractiveness (Little, Jones, Penton-Voak, Burt, & Perrett, 2002). Married couples also resemble each other in such a way that they are able to be identified as partners (Hinsz, 1989, cited in Little, & Perrett, 2002).

However, the universality of agreement on other aspects of attractiveness which we find alongside this supports the theory that some aspects of physical attraction are innately-channelled. Further evidence for the theory of innately-channelled preferences is provided by the finding that, despite all of the different modalities in which attractiveness may be rated, cues of attractiveness have the ability to provide information on the much narrower question of the genetic quality of a potential mate. Thus, one important predictor of attractiveness relates to symmetry. Comparisons of people's reactions to more and less symmetrical faces has been made by artificially manipulating images (eg Perrett *et al* 1999; Rhodes, Proffitt, Grady, & Sumich,1998; Rikowski, & Grammer, 1999), and also by twin studies, where the features of the face of each twin are measured separately to determine which is more symmetrical (Mealey, Bridgstock, & Townsend, 1999). These studies confirm the link between facial attractiveness and higher levels of symmetry. Likewise, bodily attractiveness is associated with greater symmetry Tovée, Tasker, & Benson, 2000). Voices themselves clearly cannot be more or less symmetric, but a link has been found between vocal attractiveness and measurements of bodily symmetry (Hughes *et al* 2002). Indeed, symmetrical features are found attractive by many species, as evidenced by work on mate choice in insects, birds and other mammals (Møller, & Thornhill, 1998; Møller, 1992; Møller, & Cuervo, 2003).

Cross-species attraction to symmetry is typically framed in evolutionary terms. The most symmetric ontogenetic development is achieved by a set of genes which is able to resist the various pathogens and toxins presented by the environment. Mating with the bearer of the more resistant genotype is likely to produce offspring who will benefit directly by inheriting from a genotype which is better able to withstand environmental onslaughts (and indeed symmetric ontogenetic development has been found to show considerable heritability (Livshits, & Kobyliansky, 1989), suggesting that the offspring of genetically more resistant parents are likewise themselves more resistant). The offspring of such a mating may also benefit indirectly by receiving parental care from a healthier parent. Higher levels of symmetry have been found to correlate, if fairly weakly, both with attractiveness and with better physical and mental health (Feingold, 1992; Langlois *et al* 2000; Thornhill, & Møller, 1997), including in areas, such as rural Belize, where modern medicine is less available to confound correlations with health (Waynforth, 1998)[1]. Those individuals who are innately predisposed to be attracted to such a genotype may pass on their innate predisposition to their offspring, and if those offspring are healthier, then the prevalence of the innate predisposition to symmetry within the population is likely to increase.

A second predictor of attractiveness relates to indicators of high levels of the hormones testosterone (in men) and oestrogen (in women). Thus, testosterone-enabled features such as a heavy brow ridge and large jaw add to the attractiveness of a face (Symons 1995; Thornhill, & Møller, 1997). Pitch of voice is also linked to testosterone levels. Longer and more massive vocal folds lead to lower-pitched voices, and vocal folds grow under the influence of testosterone at the later stages of puberty (review in Feinberg, Jones, Little, Burt, & Perrett, 2005). Higher testosterone levels correlate with lower-pitched voices in men (Dabbs & Malinger, 1999). Lower-pitched voices are considered more attractive (Feinberg *et al* 2005; Collins, 2000). In women, oestrogen-enabled features such as rounded upper cheeks, achieved by the laying-down of fat deposits, and large lips and breasts, increase attractiveness (Thornhill, & Grammer, 1999). The most attractive waist-to-hip ratio in women (calculated by dividing the waist circumference by hip circumference) has been found, cross-culturally, to be a figure of 0.7 (Singh, 1993; Singh, 1994, Henss 2000 etc, though see eg Gray, Heaney,

[1] Kalick, Zebrowitz, Langlois, & Johnson (1998) present dissonant evidence about the correlation between facial symmetry and health, although it is possible that modern medicine has distorted the correlation, since their subjects were United States citizens born in the 1920s. Yet the finding of a lack of a relationship between facial symmetry and health, even if consistently replicated, would not of itself disprove the suggestion that an attraction to symmetry is (or was) adaptive; it may be that there was initially a correlation, and this has now become dislocated; in this case, the ‘sexy son’ effect (discussed below) would maintain the link between symmetry and attractiveness.

& Fairhall, 2003 for criticism thereof), and this shape is achieved under the influence of oestrogen.

Zahavi's handicap principle (Zahavi, 1975, Zahavi, & Zahavi, 1997) has been used to explain how humans might be consistently attracted to indicators of high levels of testosterone and oestrogen. Testosterone acts as an immunosuppressant (Folstad, & Karter, 1992) and oestrogen gives rise to toxic by-products in the body (Service, 1998). The handicap principle, applied to mate choice, suggests that phenotypic markers of high levels of testosterone and oestrogen give honest signals about genetic quality: only the strongest and healthiest can survive the effects of these hormones. The people who are attracted to markers of high levels of these hormones are, therefore, attracted to those whose genotype makes them strong enough to deal with those handicaps. Mating with individuals with higher apparent levels of these hormones results in offspring who benefit from inheriting from the high-quality genotype. Only some individuals are able to withstand the higher levels of the hormones, and thus markers of high levels of the hormones constitutes an honest, unfakeable signal which cannot be adopted by genotypically weaker individuals. We might then expect the co-existence of markers of these hormones, and attraction thereto, to be a stable outcome, or an 'evolutionarily stable strategy', as it is expressed within the field of evolutionary game theory (Maynard Smith, 1979; Maynard Smith, 1982).

A further strand of evidence for the theory that human judgments of attractiveness are not wholly subjective or arbitrary, but rather are channelled by innate underpinnings, is the finding that female preferences change with the menstrual cycle. Females are more attracted to faces that advertise high levels of testosterone, and to the scent of more symmetric men, at the point in the menstrual cycle when conception is most likely (reviews in Gangestad, & Thornhill, 1998; Gangestad, & Thornhill, 1999; Grammer, Renninger, & Fischer, 2004; Rikowski, & Grammer, 1999). That is, female preference for the cues which indicate high-quality genetic material, and which would give rise to offspring who are best suited to survival, is strongest at the point when they are biologically most likely to conceive. We would be hard-pressed to explain such hormonally-mediated cyclic attraction to high-quality mates by reference to cultural imposition or arbitrary individual variation.

Thus, the adaptionist proposals for mate choice appear fairly robust. Indeed, adaptionist thinking can also been extended to mate choice behaviour which arises from arbitrary cultural variation, with reference to the so-called 'sexy son effect' (Weatherhead,& Robertson, 1979). If the attractiveness of a potential mate arises in part from heritable characteristics, then any offspring arising from a union with that mate may also show those characteristics. The more people who find a particular set of characteristics attractive, the better the chance that your offspring will also be attractive, by virtue of that same set of characteristics. So an innately-channelled desire to find attractive those same people who are found attractive by the rest of your society is something which might arise in a population. The Baldwin effect (Baldwin, 1896), which describes the process by which organisms evolve to become better able to learn a particular skill or behaviour, shows that there is no clear line to be drawn between learnt and innate behaviour. Both innate leanings and developmental learnings must play a part in our perception of attractiveness.[2]

1.3 Judgments of attractiveness during maturation

In sum, it seems that adults are well-equipped to pick out those mates who will help them produce healthy offspring. During maturation then, if children are to come to agree with adults about relative attractiveness, then not only must any innately-channelled dispositions become available to mate choice behaviour, but also children must learn what their culture values in a mate. In light of this body of work, one question which this present study addresses is this: if adults can be shown to form fairly consistent judgments about the relative attractiveness of a range of physical attributes, at what age do children begin to mirror these judgments? That is, at what age do people start not only tuning in on, but also giving appropriate weightings to, the many types of information relating to attractiveness which may be found in, say, the visual or vocal channels?

Connolly, Slaughter, & Mealey (2004) put forward three distinct models which we might find in relation to the longitudinal development of adult-type reactions to stimuli associated with sexual behaviour. One is that adult-like judgments are present early on in development, and apparent from an early age. Another is that adult-like judgments emerge fairly suddenly, at a

---

[2] Some of the studies mentioned above, and the question of facial and vocal attractiveness and how it fits into evolutionary theory, are also reviewed within the work submitted for assessment for the 'Animal Communications and Sociobiology' module of the MSc.

particular point in the physical or hormonal development of the child, such as at puberty. The third is that there is a gradual emergence of these judgments over time. There is some *prima facie* evidence for each of these models.

Superficial support for the first of these models arises from two sources. The first is from studies of reactions by infants to different faces. Neonates and children of only a few months in age spend longer looking at those faces which adults gauge to be more attractive (eg Samuels, Butterworth, Roberts, Grauper, & Hole, 1994), Slater *et al* 1998), irrespective of whether those faces differ from them in race, gender or age (Langlois, Ritter, Roggman, & Vaughn, 1991). However, a plausible alternative explanation for these findings is that infant preferences for attractiveness arise from the fact that attractive faces are often average-looking. The infant preference could then be explained on the basis that infants prefer faces which are composites of those faces they have seen previously, or that infants prefer faces which are more easily recognised as faces according to some kind of innate facial template (see eg Walton, & Bower, 1993, Slater *et al* 1998). Furthermore, it seems that neonates may not be using the same cues as adults in picking out particular faces. If infants, like adults, are cued by greater symmetry and greater 'averageness' in picking out attractive faces, then we would expect them to exhibit preferential staring at those faces which are symmetric rather than asymmetric, and average rather than non-average. One study, however, found the exact reverse of this pattern (Rhodes, Geddes, Jeffery, Dziurawiecz, & Clark, 2002).

The second source of support for a model whereby adult-like judgments are present from an early age arises from the suggestion that it may well be adaptive to identify genotypic and phenotypic health and strength in others, since even if those others cannot be potential mates, they may constitute social allies or, for adults, sexual competitors (Thornhill, & Grammer, 1999). They also provide indices of healthy and desirable environments, and the presence of valuable resources, all of which it is adaptive to seek out. However, we might expect to find that preferences shift at puberty as mate choice factors enter into play, and certainly the cyclic variation in female preference that coincides with the menstrual cycle cannot be manifest in pre-menarchal children.

With reference to the second model, puberty would seem, at first blush, to be a good contender for the emergence of an adult-type sexual response. Many of the sexually dimorphic traits which others find sexually attractive emerge at puberty (see eg Fitch, 2004). Sexual behaviour, precipitated by sexual attraction to others (specifically, to the secondary sexual characters of others), also emerges around puberty. This behaviour appears to be mediated by changing levels of hormones, and thus sufferers of precocious puberty, where hormone levels rise prematurely, exhibit mature sexual responses to the sexual traits of others (Thamdrup, 1961; Ehrhardt, & Meyer-Bahlburg, 1994; both cited in eg Partsch, & Sippell, 2002). We might then expect adult-like judgments to emerge at puberty along with adult-like sexual behaviour.

However, in a study on the development of preferences for the trait of waist-to-hip ratio, something which is classically associated with sexual attractiveness, it was found that the gradual-development model seemed to fit the findings best. Judgments of the most attractive waist-to-hip ratios for males and females, made by both boys and girls, moved linearly to accord with those made by adults (Connolly *et al* 2004).

The present study, then, looks specifically at the question of the emergence of adult-like judgments of facial and vocal attractiveness. There is to date a surprising dearth of studies on whether non-infant children agree with adults in judging physical attractiveness, and so this study aims to extend work on physical attractiveness by examining this area.

One predictor of an attractive voice, as identified by two separate studies (Collins, 2000; Feinberg, *et al* 2005) and mentioned above, appears to be that of lower fundamental frequency (where "fundamental frequency" is the measurable correlate of what we perceive as pitch of voice). Pitch of voice has been named a secondary sexual character cuing mate value (Feinberg *et al* 2005), and an index of mate quality (Hughes *et al* 2002). If adult-like responses to secondary sexual characters come online gradually, then we might expect to find that it affects ratings of attractiveness differently at different ages. A secondary focus of the present study therefore attempts to replicate findings on the correlation between vocal attractiveness and fundamental frequency, and also asks whether such a correlation, should it be found, exists in all of the different age-groups.

The last question addressed by this present study is whether more attractive voices belong to males with more attractive faces, and vice versa. We should expect to find that different modalities (ie the visual and the aural) supply concordant information in terms of the attractiveness of the bearer if the signals have developed under the influence of the same factor (eg high levels of testosterone), or are consistently experienced together and thus give rise to learning by association. At a different level of explanation, the adaptionist framework would suggest that we should expect to see concordant cues of mate value across different modalities if physical attraction is the sense by which we seek out high-quality mates. The sense of being physically attracted to someone should respond to those cues which denote a mate who will produce the fittest offspring. Indeed, much research, some of which is reviewed above and below, has found co-variance of attractiveness in different modalities. Furthermore, an important body of work on many different species in the discipline of animal behaviour research has found that concordant signals, giving cues to the same information ('back-up signals'), are sent through a wide range of modalities, and lead to better decoding of that signal (eg Rowe, 1999; Møller, & Pomiankowski, 1993). In the same way, in humans, concordant information may assist in the selection of high-quality and suitable mates.

Indeed, a correlation has already been found between vocal and facial attractiveness in females (Collins, & Missing, 2000). Earlier studies (Zuckerman, Miyake, & Elkin, 1995; Zuckerman, & Driver, 1989) looked for and found, at best, a very weak correlation between facial and vocal attractiveness in males, but these used entire sentences as the auditory stimuli and did not attempt to match subjects for accent, so risking confounding the narrower question of vocal attractiveness with judgments related to accent, sentence-level intonation patterns, and so on. The topic of a correlation between male facial and vocal attractiveness certainly merits further investigation, particularly in light of the various findings on attractiveness, as follows: Greater male voice attractiveness has been found to correlate with greater bodily symmetry (Hughes *et al* 2002), and in a separate study, greater body symmetry (a composite measure of the symmetry of various traits in the face, body and hands) was found to correlate with increased facial attractiveness (Hume, & Montgomerie, 2001). In addition, correlations have been found between male vocal attractiveness and sexual behaviour including number of sexual partners (Hughes, Dispenza, & Gallup, 2004), and, in a separate study, between facial attractiveness and number of short-term sexual partners (Rhodes, Simmons, & Peters, 2005). The study will also examine whether a correlation between vocal and facial attractiveness, should it exist, is to be found from the ratings of females of all ages.

1.4 The present study: outline

Females of different ages were asked to rate the relative attractiveness of the voices and faces of different males. Male ratings of females were not considered for two reasons. Firstly, as discussed above, the relationship between male facial and vocal attractiveness has not been specifically explored in an experiment when the vocal stimuli is selected to avoid unnecessary variation in intonation and accent, and so it extends existing research. Secondly and more prosaically, this limitation allowed the size of the study to be limited.

The study was not designed to replicate or expand upon the myriad other studies which have considered exactly what cues raters are using to gauge facial attractiveness. The focus of the study was instead ethological in nature, seeking to ascertain whether the mature adult, the adolescent, and the child, make different judgments on attractiveness in the real context of natural, un-manipulated visual and aural stimuli.

1.4.1 Experimental design considerations

The photographs and voice recordings were presented within a Microsoft PowerPoint 2002 (Microsoft Corporation) presentation (reproduced at 6.1 in the Appendix) to allow for portability. Two versions of the presentation were created. These were identical except that each used the photographs and recordings of just six of the males as stimuli. This division was used since the rating of twelve different males would have over-taxed particularly the younger age groups, and yet a smaller sample of just six males would have been unsatisfactory in considering possible links between facial and vocal attractiveness. These two versions are henceforth referred to as ‘Version A’ and ‘Version B’.

The male subjects who were to act as stimuli were not to have physical characteristics traditionally associated with a particular stereotype or lifestyle choice, such as tattoos, distinctive long hairstyles, or significant facial hair, in case this was something which may have created distinctions between the attractiveness ratings of different age groups by appealing to the aesthetics or culture of a particular age group or generation. Photographs of all males were shown within the presentation before the rating task began, and raters were asked to withdraw from the task if they recognised anyone. This was to reduce the risk that prior knowledge of an individual could affect attractiveness ratings, and also ethically, it was felt that the male volunteers should be assured that consideration of their relative attractiveness should not be made by anyone who knew them. It was also hoped that the initial presentation of all of the photographs should instil a degree of familiarity with the pictures, such that ratings of attractiveness would not be distorted by order of presentation.

The male subjects were selected to be of a small age-range to reduce variables. The BMI ("body mass index", an approximate measure of body fat) of subjects was calculated from self-reported height and mass to ensure a small range of variation, in case this affected the ratings of one age-group disproportionally. Current World Health Organisation recommendations list the 'ideal' BMI of an average adult as ranging from 18.5 to 24.9, and classify those with a BMI of between 25.0 and 29.9 as 'overweight' but not obese (de Onis, & Habicht, 1996). The UK population generally falls within the ideal to overweight range (Buchan, 2004), and this 'normal' range is reflected in the BMI values of the male subjects in the experiment. Although self-reports of height and weight will not be entirely precise, they were judged sufficient for the purpose of obtaining subjects with the range of BMI that was consistent with the normal population.

In order to obtain a set of comparable vocal stimuli, subjects were asked to stand, and count from one to five, pronouncing words clearly and slowly. The stimulus set only contained the counting from one to four; the number 'five' was excluded in order to avoid list-final intonation. Numbers were chosen as the spoken stimulus because they have a neutral semantic content, are easily comprehensible, and should be equally familiar to judges of all age-groups; numbers have been used as the vocal stimuli in other vocal attractiveness experiments (Hughes *et al* 2002, 2004).

Female raters were provided with a rating sheet (reproduced at 6.2 in the Appendix) to complete as they followed the presentation. Raters were given instructions along with an example of how to fill in the rating sheet. Photographs and voice recordings were presented in pairs, and for each pair, the female raters were asked to judge which one they found more attractive, and where it fell on the four-point scale represented by the phrases ‘Only a tiny bit more attractive’, A little bit more attractive’, ‘More attractive’ and ‘Much more attractive’. The word ‘attractive’ was chosen since it was deemed to have sufficient sexual connotations, but also allowed comprehension by non-sexual younger raters.

Each face was paired once with each other face. Faces were presented in the same order to each rater; the sample sizes were not large enough to guarantee an even distribution of presentation order (ie each male the same number of times presented first and on the left, first and on the right, etc) if random presentation orders had been generated, and in any case since we are comparing the ratings given by different groups of people who viewed the same presentations (rather than attempting to gauge absolute attractiveness), random presentation was deemed unnecessary. Each face was presented in a pair with every other face, such that the faces were presented either twice or three times on the left- and right-hand side, and as evenly distributed as possible. Voices were presented in the same way.

The rating scale allowed conversion to a numerical score. For each pair of faces or voices, the face or voice which was chosen as the more attractive was awarded 4 points if it were rated as ‘Much more attractive’, 3 points if ‘More attractive’, 2 points if ‘A little more attractive’ and 1 point if ‘Only a tiny bit more attractive’. Points were deducted from the score of the face or voice which was not selected from each pair, reducing their total by 4 points if the other face were assessed as ‘Much more attractive’, 3 points if ‘More attractive’ and so on. (Perhaps counterintuitively, the deductions are as necessary as the increases in score, since both faces are taken into account in selecting the appropriate phrase to describe the relative merits of each face or voice in the pair; deducting points does not simply have the effect of halving all of the total scores). This system resulted in a 41-point scale, from 20 to -20, which was treated in the analysis as an interval scale.

The pairwise comparison system was chosen following consideration of the task of rating vocal attractiveness. Whereas comparing and rating things visually is a very common task, carried out with upon any visit to the greengrocers’ or clothes shop, aural rating is much more unusual, particularly as regards very short strings of sounds. The concern, therefore, was that raters would not have a consistent internal scale for aural attractiveness to which they could refer in order to grant, say, a mark out of ten to denote relative attractiveness, and this would have led to inconsistencies in the scores awarded. It was perhaps this concern which motivated the vocal attractiveness research of Jones and DeBruine of the Universities of Aberdeen and St. Andrews (*http://www.faceresearch.org*) to use a similar four-point pairwise rating system. The same rating system was used for faces as for voices for ease of analysis and consistency of instructions, to avoid confusing raters. Furthermore, verbal labels (such as 'much more attractive') were chosen to be more accessible than a number-based scoring system to the young children participated in the task. A design where raters were asked to order faces and voices from most to least attractive was also considered, but was rejected for two reasons. Firstly, it would be difficult to recall what each voice sounded like in sufficient detail to place it in order, and potentially too time-consuming and demanding a task to listen to the voices over and again to assign ranks. Secondly, ordering the stimuli would have not shown whether any two in the final order were very different in attractiveness, or whether one was merely marginally more attractive than the other, and this would have disallowed parametric analysis of the final results.

In order to examine the question of longitudinal development of preferences, female raters were divided into three age groups, referred to as 'child', 'adolescent', and 'adult'. Adolescents were to be those who had undergone some of the behavioural and hormonal changes of puberty, but who had not reached the adult form. Hormonally, 'puberty' can be seen to begin around the age of 6 to 8, not reaching completion until around the ages of 15 to 17 (McClintock, and Herdt, 1996). However, within that period, there are two hormonal events of particular note. The first of these is adrenarche (maturation of the adrenal glands), around the age of 10. The second is gonadarche, when there is a marked increase in oestrogen, progesterone and testosterone. In girls, this is the stage when the ovaries become mature and menarche begins. Traditionally, sexual desire was thought to begin at gonadarche (Boxer, Levinson, & Petersen, 1989), but more recently, attention has focussed on adrenarche as the event which precipitates sexual desire (McClintock, and Herdt, 1996).

The division of the child and adolescent age groups was made to coincide approximately with gonadarche. As reviewed above, adult female changes in preferences are in line with the menstrual cycle; clearly such cyclical shifts cannot exist without the presence of the menstrual cycle. Asking the child participants for details about onset of menarche would have created ethical problems. However, a survey over 1000 girls of predominantly European ethnicity (Whincup, Gilg, Odoki, Taylor, & Cook, 2001) revealed that the average age of menarche was 12 years and 11 months, and that the age range had a fairly small standard deviation: menarche had taken place by the 10^{th} birthday for 0.8 % of girls, by the 11^{th} birthday for 3.6 %, and by the 12^{th} birthday for 21.7 %. To capture a group in transition between childhood and adulthood, the age range for adolescents was set at 12 to 15. The members of that group, then, would be likely to have undergone the two most significant hormonal changes, but would be a long way from full adult maturity. Adults were defined as those aged 18 and over. Children were those aged 5 to 11, with the proviso that all participants needed to demonstrate ability to take part, as set out below.

One further consideration within the experimental design was the exclusion of any female participant who may not have properly understood the task, or was otherwise unable to carry it out. To this end, raters were first asked to judge the relative attractiveness of three pairs of different images of the face of Mickey Mouse. Three of these images were taken from Gould (1980), and the other three from the work of Tagoe, Macdonald, & Caryl (unpublished). The first three images are of incarnations of the Mickey Mouse cartoon, as it was developed over the years to become increasingly 'attractive', taking on a larger relative head size, and the larger eyes and enlarged cranium associated with babyhood, which appear to be universally appealing. The remaining three images were amalgams of the original images, designed to sit between the set of original images in terms of attractiveness. The similar rating experiment of Tagoe *et al* (unpublished) found a strong degree of consistency in rating the various images for attractiveness. Females who made unusual assessments in rating these cartoon faces were to be excluded from the analysis. This task also served as a warm-up phase prior to the experiment proper, allowing familiarisation with the rating task and scale.

2. Methods

A description of the experimental design was submitted to the Ethics Committee of the Psychology Department of Edinburgh University, and approved without the requirement for any changes. A copy of the document submitted is at 6.3 of the Appendix. Since the experiment involved work with children, the researcher obtained a Basic Disclosure certificate from Disclosure Scotland.

Twelve male Caucasian native English speaking males aged 23 years 8 months to 28 years 2 months (mean age 26 years 2 months) were recruited from social contacts. Each provided his full and informed consent to the use of his photograph and voice recording in the experiment. The male subjects were photographed head-on with a neutral expression against a plain wall, using a Canon PowerShot A95 digital camera. No subjects wore any visible jewellery, and wearers of spectacles removed these for the photographs. Subjects were excluded if they had distinctive accents, were chronic smokers (more than 20 cigarettes in a week), had colds or other conditions which could affect their voice, or had ever had throat or laryngeal surgery. Subjects were recorded counting from one to five on a Sony MZ-N710 portable mini disk player using an ECM-MS907 microphone held approximately 30 centimetres from the speaker's mouth. Recording volume was adjusted during the rehearsal phase to achieve a mid-level, without distortion. BMI was calculated following self-reported height and mass, and ranged from 20.4 to 26.1, with a mean average of 23.3.

The Paint Shop Pro (Corel Corporation) computer program was used to rotate and resize the photographs such that the pupils of the eyes were all horizontal, and 310 pixels apart. Photographs were cropped so that each was presented in an oval, revealing the hairline, chin and ears, but concealing most of the hairstyle and all clothing save for a fraction of the collar. The recordings were manipulated using the program Cool Edit Pro (Adobe Systems Incorporated), such that each number was recited at a rate of one per second, by inserting copied sections of the background noise where necessary. The speech sounds were normalised for amplitude using ch_wave audio file manipulation (Edinburgh Speech Tools Library, *http://www.cstr.inf.ed.ac.uk/projects/speech_tools/manual-1.2.0/book1.htm*) in Praat (version 4.1.9, P. Boersma & D. Weenick, *http://www.praat.org*).

Female raters were recruited via social contacts from two Hampshire-based Guide packs; from a secondary school near Bristol; from visitors to a museum in Edinburgh; and from among graduate student volunteers and social networks in Edinburgh and London. Raters provided their age in years and months, but were otherwise anonymous. They were given a brief description of the task prior to their involvement, and were able to withdraw at any stage. Debriefing was provided upon completion of the task if requested. Contact details were taken of those who wished to receive a summary of the general results upon completion of the experiment. Prior to their participation, any rater aged under 16 was supplied with a description of the experiment, including contact details of the researcher and the supervisor, and required to obtain parental signature. Presentation A ran on an Aries laptop computer, and Presentation B ran on a Toshiba laptop computer. The sound recordings were played from the computer through Sennheiser high-quality headphones.

3. Results

3.1 Preliminaries

The results of 18 raters were excluded from the analysis on the basis that they did not complete the task, were unable to make a choice as to which face or voice was the more attractive for one or more pairs, or completed the form incorrectly by leaving lines blank or filling in the rating sheet in such a way that their preferences were not clear. As described in Section 1.4.1, ratings of the Mickey Mouse images were used to exclude those raters who may not have been able to carry out the task. In line with our expectations (discussed in Section 1.4.1), the very great majority of adults judged the right-hand Mickey Mouse image of rows 1 and 2 to be the more attractive, so the results of those raters whose rating sheet did not reflect this pattern were excluded from the study. The minor inconsistency in adult ratings even in the first two rows suggests that the Mickey Mouse rating test is not infallible, although it is hoped that it would remove the majority of those who did not understand the task, as well as acting as an induction phase to the exercise. Application of this criterion created the following exclusions from those who took part:

Total numbers of exclusions, per age group	Children	4
	Adolescents	3
	Adults	4
Percentage of those excluded as a proportion of their age group	Children	13 %
	Adolescents	9 %
	Adults	9 %

Table 1. Exclusions from the task. *A summary of the raters whose results were excluded from the analysis on the basis that they gave unusual ratings of the Mickey Mouse test images and therefore may not have understood the task properly.*

9 out of 47 adults judged the right-hand Mickey Mouse image of row 3 to be the more attractive, while 36 adults opted for the left-hand image. Due to the lack of consistent adult ratings, this third pair of images was excluded as a diagnostic criterion for ability to carry out the task.

Following the exclusions, the total number, and the ages, of participants in each version of the experiment was as follows:

Age group	Version	*N*	Age range	Mean age and standard deviation
Children	A	12	5 years, 8 months – 11 years, 4 months	Mean = 8 years, 7 months σ = 1 year, 9 months
	B	14	6 years, 8 months – 11 years, 11 months	Mean = 10 years, 1 month σ = 1 year, 8 months
Adolescents	A	16	12 years, 6 months – 15 years, 9 months	Mean = 13 years, 11 months σ = 1 year, 2 months
	B	15	12 years, 1 month – 15 years, 7 months	Mean = 13 years, 8 months σ =1 year, 3 months
Adults	A	24	19 years, 9 months – 34 years, 10 months	Mean = 25 years, 7 months σ = 3 years, 11 months
	B	19	20 years, 5 months – 35 years, 11 months	Mean = 26 years, 10 months σ = 4 years, 1 month

Table 2. Age of participants in each Version.

3.2 Data analysis

All analyses were conducted with SPSS version 12.0, and (unless otherwise indicated) all tests of significance were two-tailed ($\alpha = 0.05$). The raw data from the experiment is in the Appendix, at 6.4.

For purposes of comparison, the numerical scores awarded to each male (calculated as set out in 1.4.1) were first normalised by conversion to z-scores, calculated on the face ratings separately from the voice ratings.

The experiment was concerned with the main effects of, and interactions between, the ratings given to each male stimuli (faces and voices) and the age groups. In order to examine this, the data from each Version[3] were firstly analysed using a mixed-model three-factor Analysis of Variance (ANOVA) with Stimulus Modality (two levels: Face or Voice) and Stimulus Identity (six levels: six different males) as within-subjects factors and Age Group (three levels: Children, Adolescents and Adults) as the between subjects-factor. In Version B, the factor 'male' violated the assumptions of sphericity ($\chi^2 = 27.379$, p = .017), so the degrees of freedom were corrected with the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .802$).

[3] The capitalised 'Version' is used throughout to refer to Version A or B of the experiment.

This initial ANOVA revealed a significant main effect of Stimulus Identity, and a significant two-way interaction between Age Group and Stimulus Identity: that is, the different males received different attractiveness ratings when the data were collapsed across the face/voice conditions, both when the data were not separated by age group, and when the ratings of each age group were considered separately. There was also a significant two-way interaction between Stimulus Identity and Stimulus Modality: the faces and voices of the different males were rated significantly differently from each other.

Following z-score transformation, there was no significant main effect of Stimulus Modality, and no interaction between Stimulus Modality and Age Group; that is, the ratings given to the Face stimuli did not differ significantly from those given to the Voice stimuli, whether we examine the different age groups separately or not.

Full results are in Table 3.

	Version A	Version B
Age*Male*F/V	F (10, 245) = 5.061; p = < .0005	F (8.709, 187.165) = 1.255; n.s.
Age*Male	F (10, 245) = 2.909; p = .002	F (8.017, 180.391) = 3.865; p = <.0005
Male*F/V	F (5, 245) = 8.201; p = < .0005	F (4.169, 187.616) = 7.362; p = < .0005
Age*F/V	F (2, 49) =1.800; n.s.	F (2, 45) = .406; n.s.
Male	F (5, 245) = 40.504; p = < .0005	F (4.009, 180.391) = 16.384; p = < .0005
Age	F (2, 49) = 1.792; n.s.	F (2, 45) = .410; n.s.
F/V	F (1, 49) = .741; n.s.	F (1, 45) = 1.380; n.s.

Table 3. ANOVA Results. *Results of initial ANOVA showing main effects of 'Age' (3 levels: Adult, Adolescent, Child); 'Male' (6 levels: 6 different males) and 'F/V' (2 levels: Face or Voice).*

3.3 Facial and vocal attractiveness

Concordancy of judgments was calculated to check whether the stimulus set was such that it gave rise to consistent adult judgments. Without such consistency, the question of whether children made similar judgments to adults would have been misplaced. Calculations of

Kendall's coefficient of concordance (df = 5) revealed that adults made strongly concordant judgments on the facial attractiveness of the different stimuli (Version A: $W = .598$; $p = < .0005$, Version B: $W = .548$; $p = < .0005$). This replicates previous studies on concordancy of attractiveness judgments, as discussed at Section 1.2. Furthermore, the children in Version A (but not Version B), and the adolescents of both versions, showed a fair to high degree of concordancy (Adolescents: Version A: $W = .409$; $p = < .0005$, Version B: $W = .250$; $p = .002$. Children: Version A: $W = .341$; $p = .001$, Version B: $W = .131$; NS).

In contrast, Kendall's coefficient of concordance (df = 5) showed that only adults made strongly significantly concordant judgments on vocal attractiveness (Version A: $W = .560$; $p = < .0005$. Version B: $W = .364$; $< .0005$). There is significant though low concordance for the children's judgments in Version B ($W = .210$; $p = .012$), and near-significant though very low concordance for the adolescents' judgments in Version A ($W = .125$; $p = .075$). Children in Version A and adolescents in Version B do not show significant concordance.

As set out at Table 3, Version A revealed a further three-way interaction between the male stimulus, age group and face/voice factors which was not significant in Version B. However, we have hypothesised that we should see effects of age on the ratings, and furthermore the data from both Versions showed a two-way interaction between the male stimulus (when the data were collapsed across the face/voice conditions) and the age group. Therefore, we will proceed to examine the data of both Versions in order to consider whether the different age groups are rating the faces and voices of the different males significantly differently, treating any patterns which arise in Version B but not Version A with caution.

3.3.1 Facial attractiveness judgments

The Levene test for heterogeneity of variance gave a significant result for the ratings of 3 of the 6 faces in Version A and 1 of the 6 faces of Version B (Version A: Face 4: F (2, 49) = 3.529; $p = .037$. Face 5: F (2, 49) = 10.942; $p = < .0005$. Face 6: F (2, 49) = 4.411; $p = .017$. Version B: Face 11: F (2, 45) = 4.980; $p = .011$). Therefore, it was not possible to use a one-way ANOVA to determine whether the z-score ratings of the different age groups as given to the different faces differed significantly from each other. The equivalent non-parametric test, the Kruskall-Wallis (df = 2), showed that the different age groups judged Faces 1 and 6 from

Version A and Faces 8 and 9 from Version B with significant or near-significant difference (Version A: Face 1: $\chi^2 = 8.246$; $p = .016$, and Face 6: $\chi^2 = 4.798$; $p = .091$. Version B: Face 8: $\chi^2 = 8.337$; $p = .015$, and Face 9: $\chi^2 = 7.544$, $p = .023$).

In order to determine which pairs of age groups made significantly different ratings, Mann-Whitney U tests made paired comparisons of each of the age groups for these four faces. It was found that adults differed significantly (or nearly-significantly) from both children and adolescents in rating Face 1 and Face 9. Children and adolescents did not, however, differ significantly from each other in rating these two faces. The only significant differences between the age groups for the ratings given to Faces 6 and 8 was between adolescents and adults. Full results are set out in Table 4.

		Children and adolescents	Adolescents and adults	Children and adults
Face 1	*U*	75.000	126.000	63.000
	Z	-.975	-1.822	-2.718
	p	.330	.068	.007 *
Face 6	*U*	74.000	108.000	131.000
	Z	-1.022	-2.320	-.436
	p	.307	.020 *	.663
Face 8	*U*	85.500	50.000	103.000
	Z	-.851	-3.208	-1.093
	p	.395	.001 *	.274
Face 9	*U*	97.000	76.500	69.000
	Z	-.349	-2.289	-2.331
	P	.727	.022 *	.020 *

Table 4. Significant differences in facial attractiveness judgment, by age group.
** denotes significance at p = < .05.*

3.3.2 Vocal attractiveness judgments

Again, the Kruskall-Wallis test was used instead of a one-way ANOVA because the Levene test for homogeneity of variance was significant for 4 of the 6 voices in Version A and 1 of the 6 voices in Version B (Version A: Voice 1: F (2, 49) = 14.639; $p = < .0005$. Voice 2: F (2, 49) = 4.090; $p = .023$. Voice 4: F (2, 49) = 4.166; $p = .021$. Voice 6: F (2, 49) = 19.876; $p = < .0005$). The Kruskall-Wallis test (df = 2) showed that there were significant, or near-significant, differences in the judgments made by each age group for Voices 1 and 6 in Version A, and for Voices 8, 9 and 12 in Version B. (Version A: Voice 1: $\chi^2 = 19.439$; $p = < .0005$, and Voice 6: $\chi^2 = 17.995$, $p = < .0005$. Version B: Voice 8: $\chi^2 =5.964$; $p = .051$, Voice 9: $\chi^2 = 12.118$, $p = .002$, and Voice 12: $\chi^2 = 11.158$, $p = .004$.)

In order to determine which age groups were making the significantly different judgments, Mann-Whitney U tests were performed. These revealed that, for voices 1, 6 and 12, the adult age group distinguished itself from the other two age groups in terms of the ratings it awarded. For voices 8 and 9, it is the children who are making significantly different judgments from the other age groups. Full results are set out below.

		Children and adolescents	Adolescents and adults	Children and adults
Voice 1	*U*	76.000	91.000	15.000
	Z	-.928	-2.789	-4.329
	p	.353	.005 *	.000 *
Voice 6	*U*	94.000	57.000	48.000
	Z	-.093	-3.727	-3.222
	p	.926	.000 *	.001 *
Voice 8	*U*	58.000	118.000	76.000
	Z	-2.051	-.850	-2.076
	p	.040 *	.395	.038 *
Voice 9	*U*	55.500	91.000	45.000
	Z	-2.161	-1.786	-3.205
	P	.031 *	.074	.001 *
Voice 12	*U*	75.000	90.500	41.000
	Z	-1.309	-1.804	-3.351
	P	.190	.071	.001 *

Table 5. Significant differences in vocal attractiveness judgment, by age group.
** denotes significance at p = < .05.*

As we can see from Tables 4 and 5, there are four males (no.s 1, 6, 8 and 9) whose face and voice both gave rise to different ratings by the different age groups. Examination of the mean rank awarded by each age group to the faces and voices of these males (Table 6) did not, however, show any pattern across the modalities of face and voice. In other words, there was no consistent pattern where, for example, the face and voice of the same male affected child ratings in such a way that they were both given higher ratings by the children than by the other age groups.

	male 1		male 8		male 9		male 6		male 12
	Face	Voice	Face	Voice	Face	Voice	Face	Voice	Voice
Adult	1.48 **	1.17**	5.03*	3.84	1.55**	2.45	2.60*	5.48**	5.48 **
Adol-escent	1.81	2.56	3.87*	3.40	2.27	3.53	3.63*	3.31	3.31
Child	1.83	2.92	4.29	4.46**	2.32	4.43**	3.25	3.29	3.29

Table 6. Mean rank awarded by each age group to the face and voice of a subset of the male stimuli. *** denotes an age group whose ratings differed significantly from the other two age groups. * denotes pairs of age groups who differed significantly in rating a particular male, but whose ratings do not differ significantly from the third age group in rating that same male.*

3.4 Vocal attractiveness and fundamental frequency

Thus, the different age groups have been shown to make significantly different judgments in relation to a range of voices. In light of the previous work then, discussed at Section 1.3, we now turn to the question of whether there is a correlation between pitch of voice and vocal attractiveness, and whether this correlation, should it exist, is to be found for all of the age groups.

An approximate measure of the pitch of the vocal sample was determined from the fundamental frequency of the spoken word "two" using Praat (version 4.1.9, P. Boersma & D. Weenick, http://www.praat.org). Audition of each sample suggested that this word was pronounced with the most consistent intonation, and comparison of the fundamental frequency measured here with the fundamental frequency of the entire sample confirmed that it gave an appropriate index of the sample.

The measured fundamental frequency of each voice is set out at Table 7. By way of comparison, average male voice pitch ranges from 110 – 130 Hz, although a pitch falling somewhere within the range 85 – 155 Hz is not abnormal. Thus, the voices presented in both versions of the experiment contained a range of fundamental frequencies, exceeding both ends of the average male range without being abnormally high or low.

Version A		Version B	
Voice	Fundamental frequency in Hz	Voice	Fundamental frequency in Hz
2	107	9	81
4	107	8	105
1	109	10	110
5	112	11	124
3	119	7	125
6	143	12	155

Table 7. The fundamental frequency of the voice of each male stimulus.

Fundamental frequency and voice attractiveness ratings were explored with scattergraphs (reproduced at 6.5 in the Appendix) and then calculated using Pearson's coefficient of correlation, in two different ways.

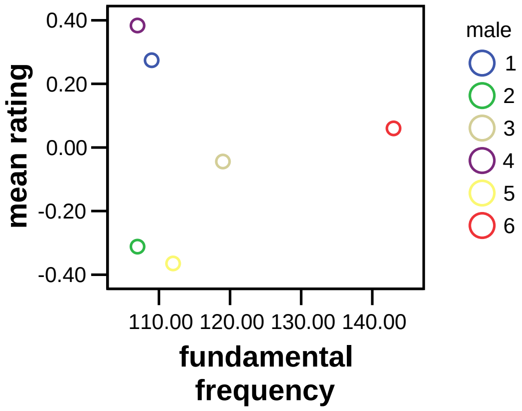
Firstly, the rating given by each rater for each male was considered as a separate token in the calculation. This analysis focuses on the raters themselves, to determine whether they identified the voice and face of the same male as being of a similar level of attractiveness. A fairly strong and highly significant relationship between a lower fundamental frequency and a more attractive voice was found for the adult age groups (Version A: $N = 144$; $r = -.492$; $p = < .0005$, Version B: $N = 114$; $r = -.596$, $p = < .0005$). In the child age group of Version B only, there is a trend in the opposite direction: more attractive voices are associated with a higher fundamental frequency ($N = 84$; $r = .247$; $p = .024$). The ratings awarded by the other age groups did not give rise to any significant correlations.

The second calculation only considered the mean rating given by each age group of each Version. Again, only the mean ratings given by the adult age groups showed a significant, or very near-significant, (and strong) correlation between lower fundamental frequency and increased vocal attractiveness (Version A: $N = 6$; $r = -.829$; $p = .041$, Version B: $N = 6$; $r = -.723$; $p = .104$ (ie .052 one-tailed).) The lack of correlation between mean rating and fundamental frequency for the children and adolescent age-groups is less surprising following the finding of a lack of concordance in their judgments of vocal attractiveness (discussed in Section 3.3), since the mean rating of a non-concordant group is a poor reflection of their opinions. Scatterplots of the vocal attractiveness rating and the fundamental frequency are set out at Figure 1.

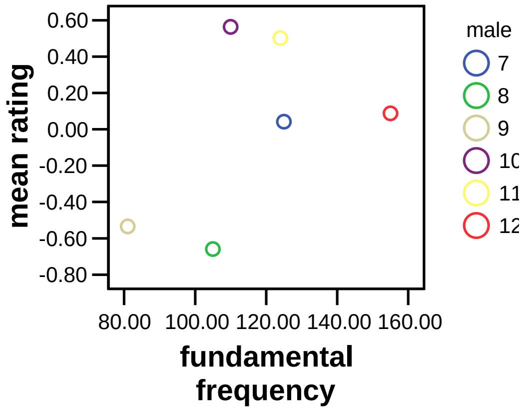
Figure 1. Fundamental frequency and mean vocal attractiveness rating, as given to each male by each age group

Children

Version A (*N* = 6; *r* = .018; NS)

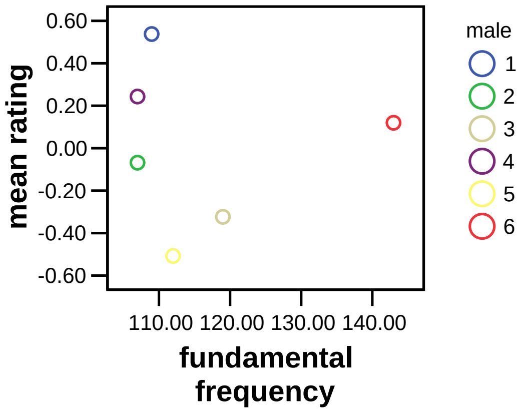


Version B (*N* = 6; *r* = .483; NS)

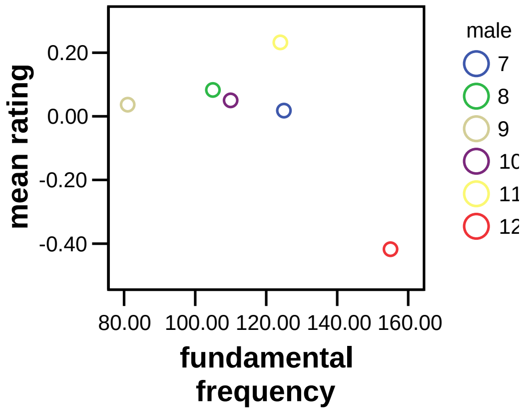


Adolescents

Version A (*N* = 6; *r* = -0.40; NS)

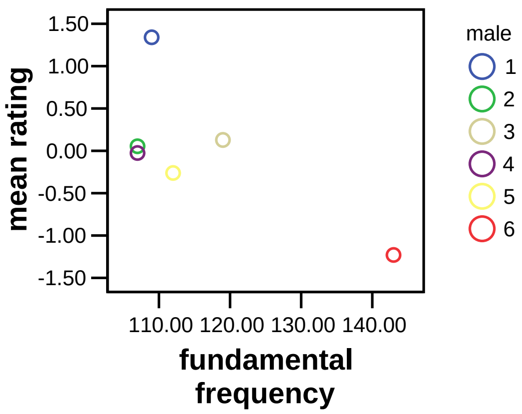


Version B (*N* = 6; *r* = -.621; NS)

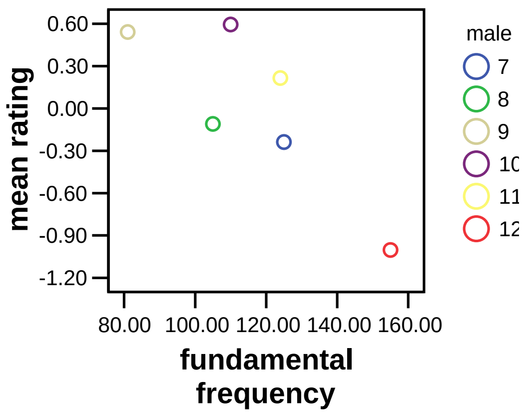


Adults

Version A (*N* = 6; *r* = -.723; *p* = .104)



Version B (*N* = 6; *r* = -.829; *p* = .041)



3.5 Co-variation of vocal and facial attractiveness

We know from the initial ANOVA that there was a main effect of the male stimulus on the ratings (see Section 3.2 and Table 3). We turn now to the question of whether male vocal and facial attractiveness co-vary. This question was addressed in two ways. Firstly, the rating awarded by each female, within each age group, was considered as a single token. This focussed on the co-variance of vocal and facial attractiveness from the point of view of the rater: were the raters picking the faces and voices of the same males as being equally attractive? In Version A, there was found to be a significant relationship between facial and vocal attractiveness as judged by each of the age groups, although the correlation is highest in the adult age group. In version B, children's rating of a male's facial attractiveness does not correlate with their rating of his vocal attractiveness. A low degree of correlation emerges in the adolescent age group, and is replicated in the adult age group, at a slightly higher level. Results are set out in full at Table 8. Scattergraphs plotting this relationship are to be found in the Appendix, at Section 6.5.

Age group	Version	*N*	Pearson correlation	Significance
Child	A	72	.266	.024
	B	84	.017	.877
Adolescent	A	96	.230	.024
	B	90	.249	.018
Adult	A	144	.333	< .0005
	B	114	.301	.001

Table 8. The correlation between the vocal and facial attractiveness rating given to each male by each rater, by age group.

In the second analysis, the mean z-score rating for the attractiveness of the face was compared to the mean z-score rating of the attractiveness of the voice of the same male. If we consider one-tailed rather than two-tailed significance on the basis that we have found evidence of a correlation between vocal and facial attractiveness above, there is a strong and significant or near-significant relationship between facial and vocal attractiveness in the ratings by children and adolescents of Version A, and in the ratings by adults in Version B. However, the finding of a correlation between facial and vocal attractiveness in the non-adult age groups is more convincing in the first calculation, because there is a lack of concordance in the judgments of either face or voice by all age groups except for the adults (discussed at Section 3.3), and the mean rating of a non-concordant group is a poor reflection of their opinions. The data is set out in full at Table 9, followed by scatterplots of the ratings in Figure 2.

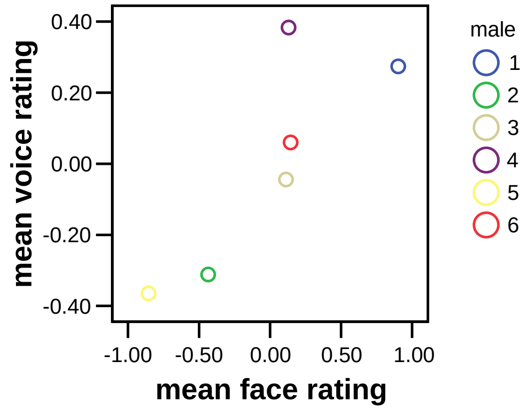
Age group	Version	Pearson correlation	Significance
Child	A	.828	.042 (ie .021 one-tailed)
	B	.159	.763
Adolescent	A	.711	.113 (ie .057 one-tailed)
	B	.508	.304
Adult	A	.461	.358
	B	.669	.146 (ie .073 one-tailed)

Table 9. The correlation between mean vocal and facial attractiveness rating for each male and each rater, by age group (N = 6).

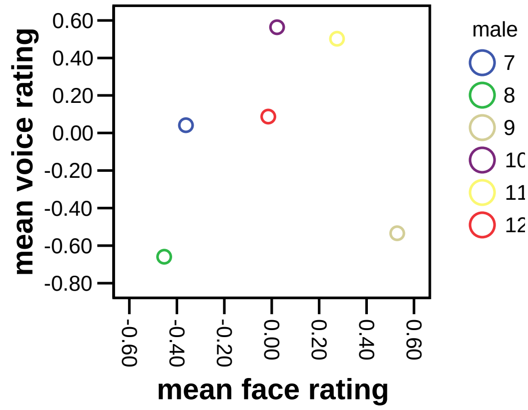
Figure 2. Mean vocal attractiveness rating and mean facial attractiveness rating (N = 6)

Children

Version A (r = .828; p = .042)

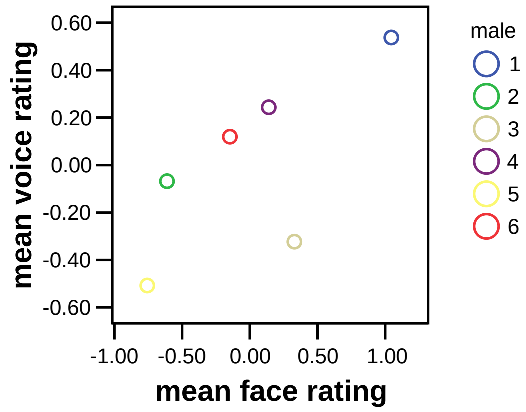


Version B (r = .159; NS)

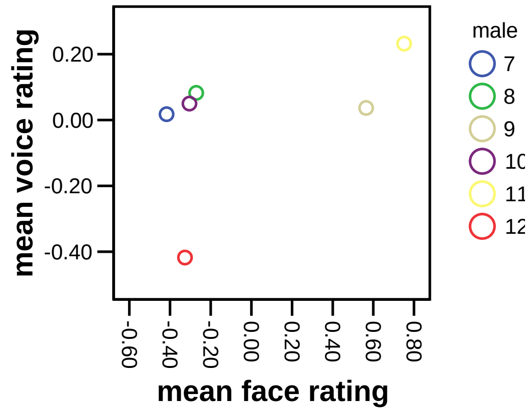


Adolescents

Version A (r = .711; p = .113)

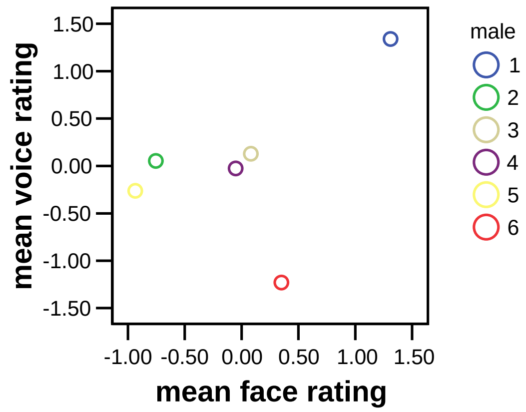


Version B (r = .508; NS)

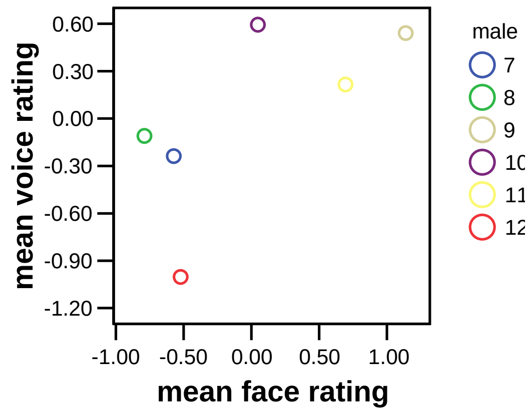


Adults

Version A (r = .461; NS)



Version B (r = .669; p = .146)



4. Discussion

4.1 Introduction

In section 1.2 three principal questions of interest were identified. These were:

- Are adult-like judgments of facial and vocal attractiveness present from an early age, or do they arise, either suddenly or gradually, during maturation?
- Is the correlation between vocal attractiveness and lower fundamental frequency replicated within this study, and if so, does it exist in the judgments made by all of the age groups?
- Is there a correlation between facial and vocal attractiveness, and if so, is it apparent from the judgments made by all of the age groups?

I now consider data from the statistical analysis with reference to these three questions.

4.2 Judgments of facial and vocal attractiveness

The first question tacked by this study is whether adult-like judgments of facial and vocal attractiveness are present from an early age, or whether they arise during maturation. We need to check that two premises are fulfilled prior to using the data from this study to answer this question. The first premise is that the stimuli (voices and faces) should be sufficiently dissimilar in attractiveness that they are able to be distinguished consistently by different adult raters. If the stimuli were very similar in attractiveness, even to the adult age group, then adults would be unlikely to make concordant judgments about their attractiveness and we would not be able to draw any conclusions about the abilities of the younger age groups. Analysis confirmed that adults were strongly and significantly concordant in their judgments of both facial and vocal attractiveness (Section 3.3), and indeed the coefficients of concordance fell within or exceeded the typical coefficient of concordance range, of 0.3 – 0.5, found between judges of facial attractiveness (Thornhill, & Gangestad, 1999b). This suggests that there is a large range of attractiveness in the stimuli and confirms that the faces and voices in each Version are valid stimuli for gaining insight into the judgments on attractiveness made by different age groups.

The second premise in the experiment is that there are no oddities about the membership of any of the different age groups in each Version, such that they are unable to present concordant judgments. The Mickey Mouse rating task (sections 1.4.1 and 3.1) was one provision designed to reduce this risk by excluding individuals who may otherwise have upset the overall ratings of their group by being unable to accomplish the task. Measures of concordance of rating (Section 3.3) show that there is a degree of concordance in all age groups and in both Versions for at least one measure (either voice or face or both), suggesting that we need have no *prima facie* concern that the make-up of any of the groups is such that it is simply unable to make concordant judgments.

The data were examined separately for facial and vocal attractiveness ratings to answer the first question identified in 4.1 above, namely whether any significant differences of judgment were made in relation to the attractiveness of each of the different stimuli by different age groups. In reference to the three possible models of development presented by Connolly *et al* (2004) and discussed in Section 1.2, there are three different scenarios which we might find.

- The first is that adult-like judgments of attractiveness are present from an early age, and if this is the case then we should not find that the different age groups make significantly different judgments.
- The second is that adult-like judgments arise fairly abruptly following the onset of puberty, and if this is the case then we should find adult-like judgments in the adolescent age group.
- The third is that adult-like judgments emerge gradually during maturation. Precise predictions are hard to make. Although a linear progression towards adult-like judgments would give rise to adolescent ratings that sit part-way between those of adults and children, we cannot assume that the the transition will progress linearly like this. Results which support this scenario will thus be harder to recognise and interpret.

We cannot assume that the model which applies to faces will also apply to voices, and so these two different modalities are considered separately.

4.2.1 Facial attractiveness

In each Version, the different age groups make significantly (or nearly significantly) different judgments from each other in relation to the attractiveness of two of the six faces (Table 4). The data are therefore inconsistent with the first model whereby adult-like judgments are present from an early age.

Nor are the data commensurate with the second model. Adult-like judgments of facial attractiveness do not appear to have arisen by the early teens, as the adolescents do not differ significantly from children in any of their judgments of the faces. One tentative interpretation might be that whatever cues the adults are using from those two faces to make their judgments are as invisible to the adolescents as they are to the children.

There is evidence both for and against the third model. On the one hand, a comparison of the measurements of concordance for each age group (Section 3.3) shows that each makes more concordant judgments than the younger age group or groups. A feature of adult judgments is that they are consistent with one another, presumably due to the availability of something equivalent to an internal scale held by each adult for judging attractiveness. Maturation requires the development of such a scale. In each Version, the coefficient of concordance measuring agreement over facial attractiveness increases by age group, so adults are more concordant than adolescents, who are, in turn, more concordant than children.[4] Thus, there appears to be some gradual movement towards something which characterises adult-like judgments. Indeed, if a scale to rate attractiveness forms gradually, and if judgments of attractiveness move from being very subjective and arbitrary to being more objective and consistent, then we would expect to see such a progression[5].

Equally, however, there is reason to question whether the third model is supported, because two faces (one in each Version) about which adults and children made similar judgments

---

[4] It is not of consequence here that, for example, the children of Version A are more concordant than those of Version B, because it could be that the stimulus set of Version B contained a smaller range of attractiveness, and thus was less distinguishable, than that of Version A. The point is that concordancy of judgment increases as the average age of the group increases, for both Version A and Version B.

[5] A potential objection to this conclusion might suggest that maturation could result in increasing concordancy of judgment about any stimulus. However, this cannot be the complete story, because judgments of vocal attractiveness are significantly and strongly concordant only within the adult age group.

were judged significantly differently by the adolescents. It seems, then, that relative to children, adolescents are moving *away* from adult-like judgments.

4.2.2 Vocal attractiveness

The ratings given to each voice were examined to determine whether the different age groups made significantly different ratings from one another. It was found that significantly (or near-significantly) different ratings were given by the different age groups for five of the twelve voices (Table 5). Further analysis revealed that for three of those voices it was only the adult age group whose ratings differed significantly; for the other two voices only the child age group did so. This pattern is consistent with the third model in 4.2 above, whereby judgments shift gradually to match those of adults during development. We might tentatively suggest that two of the voices contained cues which significantly affected the adults' and adolescents' judgments, while being invisible to the children; and that the other three voices contained cues which only significantly affected the adults' judgments, while being invisible to the adolescents and children alike. This suggestion is elaborated with reference to vocal attractiveness and fundamental frequency at section 4.3.1 below.

4.3 Vocal attractiveness and fundamental frequency

The second question tackled by this study is whether it replicates findings of a correlation between vocal attractiveness and lower fundamental frequency (Collins, 2000; Feinberg *et al* 2005; reviewed in Section 1.2), and whether such a correlation is found in the judgments made by all of the age groups.

Two separate analyses were performed, each focussing on a different aspect of this question (discussed at Section 3.4). The first took into account the rating awarded to each male by each rater. The second took into account only the mean rating awarded to each male by each age group. Both of these calculations revealed that only the adult age-groups showed a significant and indeed strong inverse trend between fundamental frequency and attractiveness: in other words, lower-pitched voices were rated more attractive and vice versa. This trend was not found in the data from the child or adolescent age group. Indeed, the first calculation showed a slight trend in one group of children to prefer *higher*-pitched voices.

Importantly, the present study has both replicated and extended the earlier findings of a correlation between male vocal attractiveness and pitch. Earlier work looked at the attractiveness of a voice stimulus reciting vowel sounds alone. Vowel sounds clearly indicate fundamental frequency and voice formant frequencies (Childers, & Wu, 1991) and thereby have the potential to convey information on the morphology of the vocal tract and the pitch of the voice. However, vowel sounds are not heard in isolation in everyday life, and are likely to downplay those aspects of speech which are created both by individual variation in articulation and also by the features of the oral cavity itself such as a tendency to lispiness. The finding that pitch of voice correlates with attractiveness despite the presence of these additional variables (in the counting stimuli used in this study) suggests that pitch may constitute a large component of what makes a voice attractive, over-riding other variation.

4.3.1 Fundamental frequency and age group

Incidental support for the suggestion that pitch of voice provides different information to different age groups is provided by consideration of those voices which the different age groups judged significantly differently. The adults set themselves apart from the other age groups in their judgments of voices 1, 6 and 12, while the children set themselves apart from the other age groups in their judgments of voices 8 and 9 (Table 5). Interesting patterns emerge when we consider these findings alongside the measured fundamental frequencies (Table 7). Voices 12 and 6 are notably higher-pitched than the other voices in the stimulus set, and they have been given a significantly lower mean rank by adults than by the other age groups (Table 6). Voices 8 and 9, in contrast, have the lowest pitch of any of the voices in the experiment, and the children have given them significantly lower mean ranks than the other age groups (Table 6). That is, the two highest-pitched voices are found significantly more unattractive by adults than by chidren and adolescents. The two lowest-pitched voices are found significantly more unattractive by children than by adults and adolescents. This is in line with an interpretation that some predilection for lower pitch has emerged by the early teens, but that a dislike of higher-pitched voices is something which develops later. This supports the conclusion drawn in 4.2.2 that we see a gradual shift towards adult-like judgments of vocal attractiveness in the adolescent age group.

4.4 Co-variance of facial and vocal attractiveness

The last question tackled by the study is whether the males with more attractive voices have more attractive faces and vice versa, and whether such a pattern, should it be found, is apparent from the judgments made by all of the age groups.

Again, two separate analyses were performed, each focussing on a slightly different aspect of the question (discussed at Section 3.5). The first took into account the rating awarded to each male by each rater (Table 8). The second took into account only the mean rating awarded to each male by each age group (Table 9).

In the first analysis, a significant correlation was found between the vocal and facial attractiveness ratings of all age groups apart from the children in Version B. In the second analysis, comparing the mean rating for facial attractiveness with the mean rating for vocal attractiveness for each age group, a strong correlation emerges from the ratings by the child and adolescent age groups of Version A and a strong and nearly-significant correlation emerges from the ratings by the adult age group of Version B. The results from the non-adult age groups of this second calculation require more cautious interpretation, because the child and adolescent age groups were either only weakly concordant or non-concordant in judging vocal attractiveness (Section 3.3), and the mean ratings of a non-concordant group are not a good reflection of the ratings provided by each member of that group. However, overall, we have good evidence that attractive voices and faces co-vary, and that this correlation is not dependent upon age of rater. This finding is commensurate with the interpretation that faces and voices provide cues to a range of concordant information, and that raters of all age groups are able to pick up on those cues.

4.5 Conclusions and extensions

We have seen that age does not appear to be an impediment to judging relative facial attractiveness consistently with one's peers, although adult-like judgments are not apparent in the younger age-groups. Indeed, there is no evidence for a linear progression towards adult-like judgments of facial attractiveness with regard to the ages represented here; it seems that different age groups are picking up on different features to form the basis of their judgments.

Vocal attractiveness judgments likewise show age-related differences. Adults appear to be cued by both high and low pitches of voice. Adolescents appear to prefer lower-pitched voices, but do not show the same aversion to higher-pitched voices that we find in adults. In contrast, higher pitch may positively influence child ratings. An explanation of the *proximate* cause of these preferences might suggest that individuals prefer the voice pitches of their peers. Only adolescents are exposed to both higher and lower pitched male voices as the voices of their male peers 'break', an event which generally occurs between the ages of 12.5 to 17.5 years, with a mean age of 13.9 years (Hagg, & Taranger, 1980). The *distal-level* explanation, however, would refer to the fact the low pitch of the male voice is a secondary sexual character providing an index of mate value by virtue of its relationship with testosterone levels. These two explanations are not inconsistent, although our focus on an evolutionary explanation requires that our dominant focus be the distal-level explanations.

In light of the differences found on voice preference by age group, further research might consider whether the specific facial cues of mate-value, such as evidence of higher levels of testosterone, also create differential responses in the older age groups. We should note, though, that the males whose voices children disliked significantly more than the other age groups, possibly because of their lower pitch, did not have faces which the children disliked more than the other age groups (Table 6). Likewise, the voices which stimulated adult dislike to a greater extent than child or adolescent preferences, possibly because of their high pitch, did not match the faces which gave rise to differential adult preferences (Table 6). Therefore, if it is testosterone (or lack thereof) which is the distal cause of adult and child dislike of a voice, then the same marker does not appear to be as salient for face judgments and does not give rise to the same significantly different judgments. Indeed, visual examination of the scatterplot of mean face and voice ratings by adults in Version A (Figure 2) suggests that their objection to the highest-pitched voice is the primary reason for the lack of correlation between vocal and facial attractiveness, which is found, in contrast, from the ratings of both of the other age groups.

As first broached in the Introduction, work on attractiveness is often interpreted in the context of adaptive strategies of mate choice. It is interesting, therefore, that the hormonal changes which appear to engender feelings of sexual attraction, and the capacity for procreation (ie gonadarche and adrenarche: see Section 1.4.1), do not appear to co-occur with the appearance of adult capabilities for judging attractiveness. There is room for further investigation as to when they do appear, and whether the emergence of a preference for, say, markers of testosterone in one modality co-occurs with the emergence of a similar preference for its expression in other modalities.

The finding of concordance between vocal and facial attractiveness from the ratings of all age groups is particularly interesting, as it provides support for the suggestion that the voice and the face alike provide an index of phenotypic or genotypic quality. The information can be used to pick out high-quality mates, or valuable social allies and benevolent environments (discussed at Section 1.3). This information is useful to individuals of all ages, and indeed it appears that those of all ages are sensitive to it. Larger samples of male stimuli would be required to tease out whether developmental shifts in preferences from childhood to adulthood tends to strengthen or weaken this correlation; or alternately whether cross-modal preferences shift in line with each other. We are familiar with the refrain that correlation does not imply causation; here, however, we should also consider the possibility that it does. That is, it is possible that a great component of the attractiveness of a voice is derived from, say, the confidence imparted to the speaker upon the realisation that he is commonly considered facially attractive. Voices provide myriad information, and listeners have been found to agree on the personality traits and emotions which they infer from the sound of the voice (Kramer, 1964; Scherer, 1979; both cited in Zuckerman, & Driver, 1989). Despite attempting to detract the effect of these traits and emotions from the voice by presenting mere vowel sounds or numbers, it may be that residual effects are still present. Attractive people have been found to have more confident behaviour (Adams, 1977; Snyder, Tanke, & Berscheid, 1977; both cited in Zuckerman, & Driver, 1989), and if this is reflected in the voice, then listeners could be basing their judgments on measures of confidence in the voice. Ideally, the voice recording should replicate the sound produced by the particular morphology of the vocal tract, independent of exterraneous factors such as accent, tension, mood (eg nervousness or otherwise of the speaker), and so on.[6] Some of these effects could

[6] The possible objection that this would have no ethological validity since speech does not exist separately from these factors can be rejected on the basis that a sufficiently large sample of voices, or hearing the same voice in a variety of circumstances, allows the individual to separate out the contributions from more intrinsic and extrinsic influences on attractiveness.

be reduced or eliminated by using vocal stimuli of a language which is not familiar to the listeners, and matching the voices for emotional content, perhaps by obtaining independent judgments of the emotional content of each voice.

Studies of the influence of the menstrual cycle on female preference have confirmed that mate choice judgments can be hormonally mediated, and, in turn, this finding of physiological influences on physical attraction has been used to support the theory of the innateness of mate choice judgments. It is necessary to enter the usual caveats regarding sample sizes and the need for replication of results, yet the present study (like that of Connolly *et al* 2004) has suggested that two key hormonal events within puberty, in contrast, do *not* immediately trigger adult-like judgments of facial and vocal attractiveness. Rather, adult-like judgments require greater maturation and even – if the suggestion that individuals prefer the voices of their peers is correct – a good element of interaction with the environment. What consequences, then, does this have for theories of an innately-derived ability to detect high-quality mates? If adult-like judgments do not emerge at the time of the key hormonal events of puberty, does this suggest that the adaptionist framework has been too eager in adopting a deflationary approach to the role of socialization?

Connolly *et al* (2004) would suggest that this may be the case: in finding that adult-like preferences for waist-to-hip ratios appear to be a function of age rather than sexual maturation, they surmise that socialization is likely to be the main driving force in developing preferences for certain body shapes. They contrast this with something which they imply is mutually exclusive with this 'socialization': that the hormones which precipitate puberty also "affect children's preferences directly, for instance by stimulating maturation of evolutionarily prewired preferences for fit mates" (p13, 2004), and call for greater work on socialization practices to understand how this works. Yet this two-part dichotomy between that which is somehow physiologically predetermined and that which is the result of 'socialization' stems from a prevalent misconception. Human adaptations have not evolved in isolation from social influences; indeed, for the purposes of natural selection, human society is just another aspect of the environment. And like all aspects of the environment, it is something which the individual can evolve to respond to. This point is nicely illustrated with reference to the imprinting behaviour of hatching chicks. Imprinting does not disprove innateness. Indeed, imprinting relies on pre-wired instincts: the chick will imprint on the first *moving* object it encounters within its critical period (see eg Bolhuis 1991). The fact that imprinting behaviour responds to external influences does not mean that imprinting behaviour is driven by socialization pressures. Likewise, the finding that human mate preferences require drawn-out maturation and experience is not inconsistent with the theory that they are guided by something which is innate. Natural selection acts upon whatever environment the individual regularly encounters. If the ‘environment’ is such, say, that females, through interaction with male peers of a similar age, consistently experience low-pitched voices at the time when they are learning what to value in a mate, then natural selection can work upon the effects of this interaction. Such is the ingeniousness of the ‘blind watchmaker’ (Dawkins, 1986).

5. References

Adams, G.R. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20,* 217-239.

Anderson, J.L., Crawford, C.B., Nadean, J., & Lindberg, T. (1992). Was the Duchess of Windsor right? A cross-cultural review of the sociobiology of ideals of female body shape. *Ethology and Sociobiology, 13,* 197–227.

Baldwin, J.M. (1896). A new factor in evolution. *American Naturalist,* 30, 441-451.

Bolhuis, J.J. (1991). Mechanisms of avian imprinting: A review. Biological Reviews, 66, 303–345.

Boxer, A., Levinson, R.A. & Petersen, A.C. (1989). Adolescent sexuality. In J. Worrell & F. Danner (Eds.), *The adolescent as decision-maker* (pp.209 – 244). San Diego: Academic Press.

Buchan, I. (2004). The public's health needs for multi-scale integration of bio-health-social research. Retrieved 10 August 2005 from *http://www.phi.man.ac.uk/Presentations/multi-scale%20bio-health-social%20research.pdf*

Childers, D.G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America, 90,* 1841-1856.

Collins, S.A. (2000). Men's voice and women's choices. *Animal Behaviour, 60,* 773-780.

Collins, S.A. & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65*, 997-1004.

Connolly, J., Slaughter, V. & Mealey, L. (2004). The development of preferences for specific body shapes. *The Journal of Sex Research, 41*(1), 5 – 15.

Cornwell, R.A., Boothroyd, L., Burt, D.M., Feinberg, D.R., Jones, B.C., Little, A.C., Pitman, R., Whiten, S., & Perrett, D.I. (2004). Concordant preferences for opposite-sex signals? Human pheromones and facial characteristics. *Proceedings of the Royal Society of London, Series B, 271,* 635-640.

Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., & Wu, C.-H. (1995). Their ideas of beauty are, on the whole, the same as ours: consistency and variability in the cross-cultural perception of female attractiveness. *Journal of Personality and Social Psychology, 68*, 261-279.

Dabbs, J.M., & Malinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27,* 801-804.

Dawkins, R. (1986). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design.* New York: W.W. Norton & Co.

de Onis, M., & Habicht, J.P. (1996). Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *American Journal of Clinical Nutrition, 64*(4), 650-658.

Ehrhardt, A.A., & Meyer-Bahlburg, H.F. (1994). Psychosocial aspects of precocious puberty. *Hormone Research 41* (supplement 2), 30-35.

Feinberg, D.R., Jones, B.C., Little, A.C., Burt, D.M., & Perrett, D.I. (2005). Manipulations of fundamental frequency and formant frequencies influence the attractiveness of human male voices. *Animal Behavior, 69,* 561-568.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304-341.

Fitch, W.T. (2004). Evolving honest communication systems: Kin selection and "mother tongues". In D.K. Oller & U. Griebel (Eds.) *The Evolution of Communication Systems: A Comparative Approach*. (pp.50-76). Cambridge, MA: MIT Press.

Folstad, I., & Karter, A.J. (1992). Parasites, bright males and the immunocompetence handicap. *American Naturalist, 139*, 603-622.

Gangestad, S.W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings of the Royal Society of London. Series B, Biological sciences, 265*(1399), 927-933.

Gangestad, S.W., & Thornhill, R. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behaviour, 20*, 175-201.

Gould, S.J. (1980). "A Biological Homage to Mickey Mouse". In S.J. Gould (Ed.) *The Panda's Thumb: More Reflections on Natural History*. (pp.95-107). New York: WW Norton.

Grammer, K., Fink, B., Møller, A.P., & Thornhill, R. (2003). Darwinian aesthetics: sexual selection and the biology of beauty. *Biological Reviews, 78,* 385-407.

Grammer, K., Renninger, L., & Fischer, B. (2004). Disco clothing, female sexual motivation, and relationship status: is she dressed to impress? *Journal of Sex Research, 41*(1), 66-74.

Gray, R.D., Heaney, M. & Fairhall, S. (2003). Evolutionary Psychology and the challenge of adaptive explanation. In K. Sterelny & J. Fitness (Eds.) *From Mating to Mentality: Evaluating Evolutionary Psychology.* (pp.247-268). London: Psychology Press.

Hagg, U., & Taranger, J. (1980). Menarche and voice change as indicators of the pubertal growth spurt. *Acta Odontologica Scandinavica, 38*(3), 179-186.

Henss, R. (2000). Waist-to-hip ratio and female attractiveness. Evidence from photographic stimuli and methodological considerations. *Personality and Individual Differences, 28*(3), 501-513.

Hinsz, V.B. (1989). Facial resemblance in engaged and married couples. *Journal of Social and Personal Relationships, 6,* 223-242.

Hughes, S.M., Dispenza, F., & Gallup, G.G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25*, 295-304.

Hughes, S.M., Harrison, M.A., & Gallup, G.G. (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behaviour, 23*, 173-180.

Hume, D.K., & Montgomerie, R. (2001). Facial attractiveness signals different aspects of “quality” in women and men. *Evolution and Human Behavior, 22,* 93-112.

Kalick, S.M., Zebrowitz, L.A., Langlois, J.H., & Johnson, R.M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science, 9*(1), 8-13.

Kramer, E. (1964). Personality stereotypes in voice: A reconsideration of the data *The Journal of Social Psychology, 62,* 247-251.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M. & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126,* 390-423.

Langlois, J.H., Ritter, J.M., Roggman, L.A. & Vaughn, L.S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology, 27*, 79-84

Little, A.C., Jones, B.C., Penton-Voak, I.S., Burt, D.M., & Perrett, D.I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shapes. *Proceedings of the Royal Society of London B, 269,* 39-44.

Little, A.C., & Perrett, D.I. (2002). Putting beauty back in the eye of the beholder. *The Psychologist, 15*(1), 28-32.

Livshits, G., & Kobyliansky, E. (1989). Study of genetic variance in the fluctuating asymmetry of anthropometrical traits. *Annals of Human Biology, 116,* 121-129.

Maynard Smith, J. (1979). Game theory and the evolution of behaviour. *Proceedings of the Royal Society, London, B 205*, 475-488.

Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

McClintock, M.K., & Herdt, G. (1996). Rethinking puberty: the development of sexual attraction. *Current Directions in Psychological Science, 5*(6), 178-183.

Mealey, L., Bridgstock, R., & Townsend, G.C. (1999). Symmetry and perceived facial attractiveness: a monozygotic co-twin comparison. *Journal of Personality and Social Psychology, 76*, 157-165.

Møller, A.P. (1992). Female swallow preference for symmetrical male sexual ornaments. *Nature, 357,* 238-240.

Møller, A.P., & Cuervo, J.J. (2003). Asymmetry, size and sexual selection: factors affecting heterogeneity in relationships between asymmetry and sexual selection. In: M. Polak M (Ed.) *Developmental instability: causes and consequences.* (pp.112-135). Oxford University Press, New York.

Møller, A.P., & Pomiankowski, A. (1993). Why have birds got multiple sexual ornaments? *Behavioral Ecology and Sociobiology, 32,* 167-176.

Møller, A.P., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: A meta-analysis. *American Naturalist, 151,* 174-192.

Partsch, C.-J., & Sippell, W.G. (2002). Treatment of central precocious puberty. *Best Practice & Research Clinical Endocrinology and Metabolism 16*(1), 165-189.

Perrett, D.I., Burt, M.D., Penton-Voak, I.S., Lee, K.J., Rowland, D.A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behaviour, 20*, 295-307.

Perrett, D.I., Penton-Voak, I.S., Little, A.C., Tiddeman, B.P., Burt, D.M., Schmidt, N., Oxley, R., Kinloch, N., Barrett, L. (2002) Facial attractiveness judgements reflect learning of parental age characteristics. *Proceedings of the Royal Society of London B, 269,* 873-880.

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings of the Royal Society London, Series B: Biological Sciences, 266*, 869-874.

Rhodes, G., Geddes, K., Jeffery, L., Dziurawiecz, S., & Clark, A. (2002). Are average and symmetric faces attractive to infants? Discrimination and looking preferences. *Perception, 31*, 315 – 321.

Rhodes, G., Proffitt, F., Grady, J., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin and Review, 5*, 659-669.

Rhodes, G., Simmons, L.W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior, 26,* 186-201.

Rhodes, G., Zebrowitz, L.A., Clark, A., Kalick, S.M., Hightower, A. & McKay, R. (2001). Do facial averageness and symmetry signal health? *Evolution and Human Behavior, 22,* 31-46.

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry and attractiveness. *Proceedings of the Royal Society of London, Series B, 266*(1422), 869-874.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour, 58*, 921-931.

Samuels, C.A., Butterworth, G., Roberts, T., Grauper, L. & Hole, G. (1994) Facial aesthetics: Babies prefer attractiveness to symmetry. *Perception, 23*, 823-831.

Scherer, K.R. (1979). Voice and speech correlates of perceived social influence. In H. Giles & R.N. St Clair (Eds.) *Language and social psychology.* (pp. 88 – 120). London: Arnold.

Service, R. New role of estrogen in cancer. *Science,* 279:1631–1632, 1998

Singh, D. (1993). Adaptive significance of female physical attractiveness: role of waist-to-hip ratio. *Journal of Personality and Social Psychology, 59,* 112–120.

Singh, D. (1994). Is thin really beautiful and good? Relationship between waist-to-hip ratio (WHR) and female attractiveness. *Personality and Individual Differences, 16,* 123-132.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development, 21,* 345-354.

Snyder, M., Tanke, E., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656-666.

Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In P.R. Abramson & S.D. Pinkerton (Eds.), *Sexual Nature, Sexual Culture.* Chicago: UCP.

Tagoe, J.D., Macdonald, N. & Caryl, P.G. (unpublished). Female attractiveness: the role of waist-hip ratio (WHR) and body mass index (BMI) before and after puberty. Unpublished poster, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom.

Thamdrup, E. (1961). *Precocious Sexual Development, A Clinical Study of 100 Children.* Kopenhagen: Munksgaard.

Thornhill, R., & Gangestad, S.W. (1999a). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior, 20,* 175-201.

Thornhill, R., & Gangestad, S.W. (1999b). Facial attractiveness. *Trends in Cognitive Science, 3*(12), 452-459.

Thornhill, R., & Grammer, K. (1999). The body and face of woman: one ornament that signals quality. *Evolution and Human Behavior, 20*, 105-120.

Thornhill, R., & Møller, A.P. (1997). Developmental stability, disease and medicine. *Biological Review, 72*, 497-548.

Tovée, M.J., Tasker, K., & Benson, P.J. (2000). Is symmetry a visual cue to attractiveness in the human female body? *Evolution and Human Behavior, 21,* 191-200.

Walton, G.E., & Bower, T.G.R. (1993). Newborns form "prototypes" in less than 1 minute. *Psychological Science, 4,* 203-205.

Waynforth, D. (1998). Fluctuating asymmetry and human male life history traits in rural Belize. *Proceedings of the Royals Society of London B, 265,* 849-853.

Weatherhead, P.J., & Robertson, R.J. (1979). Offspring quality and the polygyny threshold: the sexy son hypothesis *American Naturalist, 113*, 349-356.

Whincup, P.H., Gilg, J.A., Odoki, K., Taylor, S.J.C., & Cook, D.G. (2001). Age of menarche in contemporary British teenagers: survey of girls born between 1982 and 1986. *British Medical Journal, 322,* 1095-1096.

Zahavi, A. (1975). Mate selection – a selection for a handicap. *Journal of Theoretical Biology, 53*, 205-214.

Zahavi, A. and Zahavi, A., 1997, *The handicap principle: A missing piece of Darwin's puzzle*. Oxford: Oxford University Press

Zebrowitz, L.A., & Rhodes, G. (2000). Nature let a hundred flowers bloom: the multiple ways and wherefores of attractiveness. In: G.Rhodes & L.A.Zebrowitz (Eds.), *Advances in Visual Cognition, vol.1, Facial attractiveness.* Westport, CT: Ablex.

Zuckerman, M. & Driver, R. (1989). What sounds beautiful is good: the vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13,* 67-82.

Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype: replication and elaboration. *Journal of Nonverbal Behavior, 14,* 97-112.

Zuckerman, M., Miyake, K., & Elkin, C.S. (1990). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality, 29,* 253-272.