Accuracy of Facial Approximation:

Studies in measurement, prediction, and "recognizability" of human face anatomy

Carl Nathaniel Stephan

This thesis is submitted to the Department of Anatomical Sciences,
The University of Adelaide, in fulfillment of the requirements for the degree of
Doctor of Philosophy in Biological Anthropology.
2003

THE UNIVERSITY
OF ADELAIDE
AUSTRALIA

Amendments to Thesis

Accuracy of Facial Approximation: Studies in measurement, prediction, and
"recognizability" of human face anatomy
by Carl N. Stephan

Throughout the thesis:
"Iscan" should be replaced with "İşcan"
"Stadmuller" should replace incorrect spellings, e.g., "Studmuller"
Referenced page numbers should be consistently placed inside parentheses

P 14, line 3, comma should be inserted after "age"
P 15, first paragraph should be correctly aligned
P 15, line 3, "homo" should read "*Homo*"
P 18, line 5, "Y. Iscan and W. Krogman" should read "M.Y. İşcan and W.M. Krogman"
P 18, line 5, Krogman and İşcan is not an edited text and therefore these authors should be cited as sole authors.
P 19, line 20, insert sentence: "Note here that 2 dimensional drawing techniques are faster and less expensive than three-dimensional sculpting or computerized methods (see Krogman and İşcan 1986).
P 19, lines 9,11, & 12, "(i)" "(ii)" and "(iii)" should be replaced with "(a)", "(b)" and "(c)" respectively
P 23, line 7, "soft- to hard" should be replaced with "soft-to-hard"
P 23, line 19, sentence beginning "Consequently, it seems that…." should read: "Consequently, it seems that combination methods of facial approximation, while not optimal, are the only way forward."
P 28, line 22, "eventhough" should read "even though" and "p.33" should read "p. 33"
P 28, last line, "Mancehester" should be spelt "Manchester"
P 29, line 4, "acknowedge" should be spelt "acknowledge"
P 28 & 30 "et al." should be replaced with "*et al.*"
P 36, line 16, "recognizabiltiy" should be spelt "recognizability"
P 41, line 19, sentence ending "…type and storage)." should read "…type and storage); and relatively high cost."
P 107, Table 17, table key, as indicated below, should be added:
* indicates articles cited in Drews (1957)
† indicates articles cited in Knudtzon (1949)
P 114, initial of "E." in "E. Craig" should be removed throughout text

Behrents (1985) reference should include the publisher (The University of Michigan)
Krogman (1962) reference and Krogman and İşcan (1986) reference should have correct city name ("Springfield" should replace "Illinois")
Stadmuller (1922 and 1925) references - "Anthopologie" should be spelt "Anthropologie"

Declaration

This dissertation represents my own work that has been, in part, conducted with the advice, supervision, and support of Prof. Maciej Henneberg, Prof. John Clement and others who assisted with various projects. It is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Anthropology. It has not been submitted in this or in similar form for the requirements of any other degree at any other university. Any mistakes are mine.

I give consent to this copy of my thesis, when deposited in the University Library at the appropriate time, being available for loan and photocopying.

NAME: DATE: 30/4/03

Carl Nathaniel Stephan

Acknowledgments

Many people have provided assistance, advice and inspiration throughout the years of my PhD study. I sincerely thank them all for their help. In particular: my supervisors Prof. Maciej Henneberg and Prof. John Clement who provided me with endless support and encouragement; The School of Dental Science, The University of Melbourne who permitted access to standardized craniofacial photography equipment; Prof. David Perrett, Dr. Bernard Tiddeman and Dr. Ian Penton-Voak who generously provided advice and use of computer software required for much of this study; The Prince Alfred Hotel who assisted in compensating average face participants for their time; Prof. Wayne Sampson for advice and research support; Ronn Taylor for his invaluable mentoring on facial approximation and great discussions; Dr. Jane Taylor for technical assistance and encouragement; my parents, Christine and Colin, for support of my academic pursuits in everyway, especially by tolerating me when I was (frequently) at home in times of poor finance; Ronn (again) and Chris and Clay Taylor who welcomed me into their home on numerous occasions during research periods in Melbourne; Zac and Lesley Stephan who helped support me when traveling scholarships fell short; Dr. Jeff Trahair, Dr. Mary Katsikitis, and Rachel Norris who all provided additional inspiration, valued advice, and discussion; and editors and reviewers of the *American Journal of Physical Anthropology*, *Forensic Science International*, *Kluwer Academic Publishers* and the *Journal of Forensic Sciences* for comments and advice on research extracts from this thesis. Last but certainly not least, I send a very special thanks to the many volunteers who took part in this study, especially those who consented to publication of their photographs - your generosity and contributions were priceless.

Contents

Figures

Tables

Publications Directly Resulting From This PhD

Stephan C, Clement J, Owen C, Dobrostanski T, Owen A. A new rig for standardized craniofacial photography put to the test. Manuscript in 2nd review with Plastic and Reconstructive Surgery.

Stephan CN, Henneberg M, and Sampson W. Predicting nose projection and pronasale position in facial approximation: A test of published methods and the proposal of new guidelines. American Journal of Phsycial Anthropology, In press. Accepted 21-2-2003.

Stephan CN, and Henneberg M. Predicting mouth width from canine width – a 75% rule [technical note]. Journal of Forensic Sciences, In press. Accepted 14-2-2003.

Stephan CN. Facial approximation: An evaluation of mouth width determination. American Journal of Physical Anthropology 2003; 121 (1): 48-57.

Stephan CN. Position of superciliare in relation to the lateral iris: Testing a suggested facial approximation guideline. Forensic Science International 2002; 130 (1): 29-33.

Stephan CN. Facial approximation: Globe projection guideline falsified by exophthalmometry literature. Journal of Forensic Sciences 2002; 47 (4): 730-735.

A Version of this paper was presented at the 16th International Symposium on the Forensic Sciences 2002. Canberra, Australia.

Stephan CN. Do resemblance ratings measure the accuracy of facial approximations? Journal of Forensic Sciences 2002; 47 (2): 239-243.

A *version of this paper was presented at the International Association of Craniofacial Identification 2000. FBI headquaters, Washington D.C., USA..*

Parts of the chapter entitled "Ceiling Limits of Recognition From Two-Dimensional Facial Approximations" have also been presented at the 54th Annual meeting of the American Academy of Forensic Sciences 2002. Atlanta, Georgia, USA.

"…no-one knows quite how [facial approximation practitioners] do it. Maybe [they] have messages beamed to them from outer space"

Rubin (1998, pp. 67). *Murder He Sculpted.*

Abstract

Methods of facial approximation, whilst being controversial, have considerable forensic significance because they are one of the few anthropological techniques that have the consistent potential to obtain specific and purposeful individual identifications from skeletal remains. Fundamental to the accuracy of facial approximation is the knowledge of anatomical facial structures, their relationships and their variabilities. Here, commonly used methods for predicting soft tissue anatomy of major face features (eyes, nose, mouth and eyebrows) from the skull were tested and improved. New guidelines which better predict mouth width are presented (improving accuracy by up to 13mm on average), as are new regression equations that better predict pronasale position (improving accuracy by up to 11mm on average), average expohthalmometry measures which better predict eyeball projection (improving accuracy by about 4mm on average), and new guidelines that better predict superciliare position (improving accuracy by up to 5mm on average). Since all previously untested guidelines tested here could be considerably improved, it seems likely that many other previously untested soft tissue prediction guidelines could also have their accuracy increased in the future.

Enhanced computer graphic techniques were also used to generate standardized average human face anatomies for an Australian sample aged from 18 to 34 years. These average faces were used to remove aspects of subjectivity present in traditional facial approximation methods. By warping average faces to exact face shapes, best-case scenario (or ceiling level) true positive recognition rates were established for two-dimensional facial approximations (≈43%). By warping average faces to fit skulls according to the improved soft tissue prediction guidelines and other commonly used guidelines, and testing facial approximation

recognition by using face pools, it was found that these facial approximations were recognized well below ceiling rates and were not recognized at rates statistically above that expected by chance (5%). These findings indicate that further research is needed to empirically determine other anatomical relationships that can be used for predicting soft tissue anatomy from the skull before any current facial approximation methods can be expected to generate faces that can be specifically and "reliably" (i.e., ≈43%) recognized as target individuals.

Introduction

Biological Anthropology and Facial Approximation

Biological (physical) anthropology can be defined as the study of the physical attributes of the human body. Hence, one of the core components of biological anthropology is an understanding of the variation that exists in human biological characters.

Since soft tissues decompose rather quickly, skeletal remains are frequently the only remaining evidence of human bodies and, therefore, skeletal analysis plays a significant role in biological anthropology. Knowledge of the variability of human osteology is, consequentially, essential to the general understanding of human form. It is also necessary for accurately estimating an individual's biological profile, i.e., sex, age, height, population of origin, pathology, handedness, occupation and even soft tissue facial appearance among other characteristics (Iscan and Kennedy, 1989; Krogman and Iscan, 1986; Reichs, 1998). Biological profile estimation is useful in biological anthropology, not only for academic pursuits like research in human macro- and micro-evolution (Henneberg *et al.*, 2002), but also for forensic science where it is applied to aid the identification of skeletal remains (Iscan and Kennedy, 1989; Krogman and Iscan, 1986; Reichs, 1998; Stewart, 1970; Stewart, 1979a).

One component of biological profile estimation, soft tissue facial prediction from the skull or facial approximation (Fig. 1), the subject of this thesis, is a significant anthropological technique in a forensic context because it is one of the few anthropological methods currently being used that has the consistent potential to obtain specific and purposeful individual identification in an absence of non-skeletal information (another anthropological method that allows specific and purposeful individual identifcation on a "consistent" basis is radiographic comparison of frontal sinus shapes). Unlike the face, which is practically unique to all individuals (perhaps excluding monozygotic twins), other biological characters estimated using anthropological techniques, like sex, age population of origin etc., are shared by many individuals, and consequently cannot be reliably used for specific personal identification, although they may be used to narrow investigations by excluding people to whom the remains do not belong, which is also useful (Gill, 1990; Rhine, 1990b; Sauer, 1992).

The importance of a good understanding of human biological diversity in biological anthropology is especially apparent in facial approximation methods because the process relies on a detailed understanding of the human facial form at many levels. For example, to build a face from a skull knowledge of hard tissue morphology and its variation are required, as well as an understanding of soft tissue morphology and its variation, and perhaps most importantly, an understanding of the relationship between both these tissues and knowledge of how their relationships vary between individuals.

The significance of an accurate understanding of human variation is also amplified in forensic applications of facial approximation since the ramifications of inaccurate or wrong predictions of biological characters of the face can be significant. For example, victim identification is often the ultimate factor contributing to the apprehension of criminals (Brues, 1992). If wrong predictions are made with respect to identifying a victim, the criminal (often murderers in the case of facial approximation) may not be caught and may have the opportunity to re-offend. Additionally the victims families may be placed in a situation were their grieving is prolonged or becomes more intense, which is not favorable. In contrast, wrong predictions made by facial approximation practitioners in an academic environment (i.e., for the visualization of human ancestors, Fig. 1) caused by inaccurate knowledge of human variation, while still important (see e.g., Montagu (1947)), may be seen to be not so significant because it results in inaccurate perceptions rather than the opportunity for additional crime or exaggerated grieving.

The use of facial approximation techniques on ancestral skulls of modern homo will be discussed little further in this thesis since these facial approximations are fundamentally flawed, as has been previously reported (Montagu, 1947). The decomposition of the soft tissue parts of paleoanthropological beings makes it impossible for the detail of their actual soft tissue face morphology and variability to be known, as well as the variability of the relationship between the hard and the soft tissue. As a result, the faces of earlier human ancestors cannot be objectively constructed or tested. Attempts based on modern ape morphologies (and variabilities) are likely to be heavily biased, grossly inaccurate, and invalid because modern apes are likely to be far removed from earlier ancestral forms as a result of secular trends and evolutionary forces (Montagu, 1947).

(a) **(b)**

Figure 1: Examples of various types of facial approximations

(a) Forensic three-dimensional clay facial approximation (left) and target individual (right); **(b)** Facial approximation of a Neanderthal, image reproduced from Montagu (1947).

The History of Forensic Facial Approximation

Facial approximation has most commonly been known as "facial reconstruction" in the past (Tyrrell *et al.*, 1997), but it has also been referred to as "facial reproduction" (Rhine, 1990a), "facial reconstitution" (Suzuki, 1973), "facial restoration" (Farrar, 1977) and "forensic sculpture" (Gatliff, 1984). It has been often been suggested that many of these other names do not adequately describe the procedure since they either imply false accuracy and/or have already been used to describe other methods (George, 1987; Rhine, 1990a; Stephan and Henneberg, 2001), however it has not been consistently agreed which names fall under these categories. The most appropriate name appears to be "facial approximation", as originally described by George (1987), since it indicates an inexact face building technique and because this term has not been used to describe other methods. In contrast, the commonly used name "facial reconstruction" (falsely) implies a high degree of accuracy (George, 1987), and is already used to describe the process of reassembling bony skull fragments (Rhine, 1990a) and methods of facial medical surgery (Converse, 1977). Many agree that "facial approximation" is the most appropriate term (George, 1987; Stephan and Henneberg, 2001; Taylor, 2001a) but some have not used it because it is not in general circulation (Taylor, 2001a). This is not favorable since false impressions of methodological accuracy are sustained and further promoted (George, 1987; Rhine, 1990a). Therefore, "facial approximation" will be used throughout this thesis.

Techniques of facial approximation have been used for a considerable amount of time as indicated by the archeological finds of plastered faces on skulls in Jericho dating back to about 5500 BC (Prag and Neave, 1997). As previously mentioned, they have also been more recently employed by biological anthropologists (and others, like artists) in order to visualize the appearance of human ancestors, but of particular interest in this thesis, is their application to forensic science to help identify skeletal remains. Facial approximation is used in a forensic context with the specific aim of producing a face that can be purposefully recognized as the target individual (person to whom the skull belongs) when the facial approximation is advertised in the media (Prag and Neave, 1997; Taylor, 2001a).

The first reports of "systematic" facial approximation attempts are from Germany in the late 1800's by Welcker (1883) and His (1895a; 1895b), who used average soft tissue measures to build facial approximations, some for the purpose of confirming individual identification (His, 1895a; His, 1895b). Since being initially developed in Germany (His, 1895a; His, 1895b; von Eggeling, 1913; Welcker, 1883) forensic facial approximation methods have become widely used, as reflected by the international origin of many forensic facial approximation publications, e.g., papers have originated from: Australia (Stephan and Henneberg, 2001; Taylor and Angel, 1998); Britain (Evison, 2001; Prag and Neave, 1997; Tyrrell *et al.*, 1997; Wilkinson, 2002); Germany (Helmer, 1984; Helmer *et al.*, 1993); Japan (Suzuki, 1948; Suzuki, 1973); New Zealand (Stoney and Koelmeyer, 1999); Russia (Gerasimov, 1971; Lebedinskaya *et al.*, 1993); Scotland (Vanezis *et al.*, 1989; Vanezis *et al.*, 2000); South Africa (Aulsebrook *et al.*, 1996; van Rensburg, 1993); and the United States of America (Gatliff, 1984; Gatliff and Snow, 1979; Gatliff and Taylor, 2001; George, 1987; Krogman, 1962; Rubin, 1998; Taylor, 2001a; Ubelaker and O'Donnell, 1992).

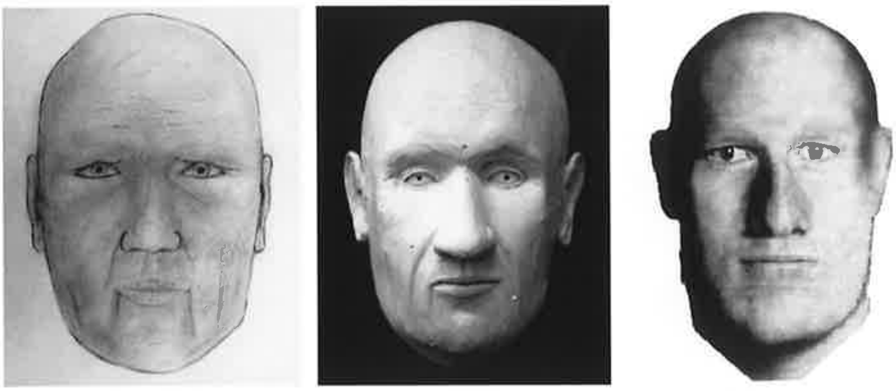
The Russian development of the method, for which M. Gerasimov (1971) is famous, did not employ average soft tissue depths like the original technique developed in Germany. M. Gerasimov's technique (known as the Russian technique (Prag and Neave, 1997)) relied primarily on building the facial musculature. During the time M. Gerasimov was practicing, the average soft tissue depth method also became popular in the U.S.A. primarily due to two practitioners: a physical anthropologist, W. Krogman, who used and promoted the technique from about the mid 1940's (Krogman, 1946; Krogman, 1962; Krogman and Iscan, 1986), and a forensic artist, B. Gatliff, who is probably the most renowned U.S.A. facial approximation practitioner, becoming well known since the 1970's as reflected by her publications (Gatliff, 1984; Gatliff and Snow, 1979; Gatliff and Taylor, 2001; Snow *et al.*, 1970). While the average soft tissue method was originally employed in Germany, it has become known as the American technique (Prag and Neave, 1997). At about the same time as B. Gatliff was working on facial approximation, R. Neave, a medical artist in Britain, was developing a new version of the technique by combining both the American and Russian approaches. Although B. Gatliff and R. Neave are perhaps the most famous facial approximation practitioners in recent times, it seems that the highest academically acclaimed facial approximation experts during similar time frames are probably Y. Iscan and W. Krogman who co-edited one of the main text on the topic (Krogman and Iscan, 1986) and who have contributed to the facial approximation literature with many other publications that have become landmark references (e.g., Iscan and Helmer, 1993).

Facial approximation has been a controversial technique for most of its existence because it seems faces built from skulls are not always (maybe infrequently) accurate (Brues, 1958; Diedrich, 1926; Haglund and Reay, 1991; Montagu, 1947; Stadmuller, 1922; Stadmuller, 1925; Stephan and Henneberg, 2001; Suk, 1935; von Eggeling, 1913). von Eggeling (1913) first criticized the method in a formal publication in 1913, about 40 years after the first attempts at building faces using average soft tissue depths (His, 1895a; His, 1895b). Since then this controversy has not been resolved and there has been a steady flux of critical papers since von Eggeling's 1913 article (e.g., Brues, 1958; Diedrich, 1926; Haglund and Reay, 1991; Montagu, 1947; Stadmuller, 1922; Stadmuller, 1925; Stephan and Henneberg, 2001; Suk, 1935). Although everyone in the field seems to recognize that facial approximation methods will never achieve an exact likeness (Vanezis *et al.*, 1989), there are those who believe that the method still achieves its goal (specific and purposeful facial recognition) and those who are not so convinced. Despite the original criticisms by von Eggeling (1913), and those by Suk (1935) and Montagu (1947) after which facial approximation methods reportedly fell into disrepute (Prag and Neave, 1997; Taylor, 2001a), the methods have probably become, overall, more popular with time, at least with laypersons.

Forensic Facial Approximation Methods

Currently, techniques of forensic facial approximation are regarded as "last resort" methods that are used to produce tentative identifications (Caldwell, 1986; Gatliff, 1984) when other identification methods, like DNA comparisons, dental comparisons and fingerprint comparisons, are not possible. These tentative identifications are significant because they potentially allow for positive identifications to be made via other (more reliable) methods (like DNA comparison), and they also often generate new leads in difficult cases. Forensic facial approximations are usually constructed by a forensic anthropologist or a forensic artist, or by both working in collaboration with one another on the same skull.

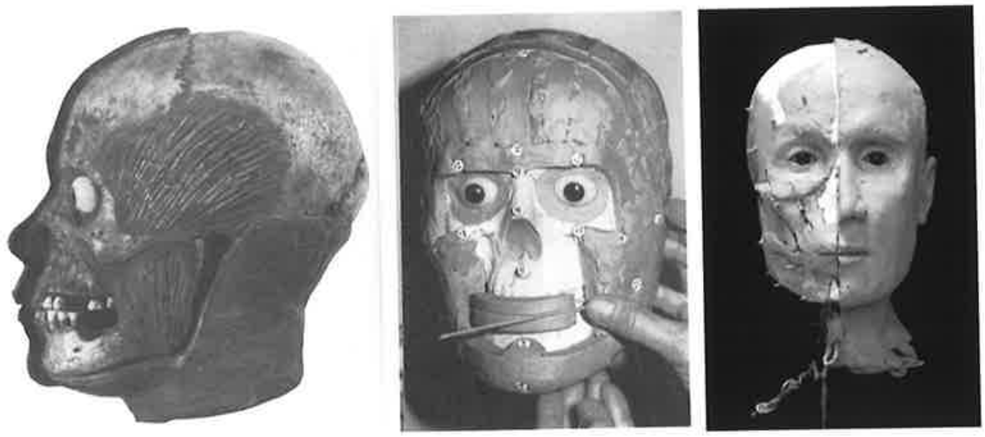
Facial approximation is usually accomplished by one of three approaches: (i) drawing on acetate paper over an image of the skull (two-dimensional approximation) (Taylor, 2001a); (ii) sculpting clay over the skull or a skull cast (three-dimensional approximation) (Gatliff, 1984; Gerasimov, 1971; Prag and Neave, 1997); or (iii) using computer imaging techniques which usually involve warping a contour surface map of a face to "fit" the skull (two/three-dimensional approximation) (Evison, 2001; Perper *et al.*, 1988; Vanezis *et al.*, 2000). Figure 2 shows examples of each of these approaches. Traditionally drawing and clay sculpting techniques have been the most frequently used (Taylor, 2001a). Some have indicated that three-dimensional clay facial approximations are more popular than drawing techniques, which are rarely used (Tyrrell *et al.*, 1997), but this claim may only be correct in a very general sense since there are some practitioners who specialize in, and frequently use, drawing techniques (e.g., Taylor, 2001a). Additionally, while computer generated techniques may be used primarily at some institutions, e.g., The University of Glasgow (Vanezis *et al.*, 2000), and The University of Sheffield (Evison and Green, 1999; Tyrrell *et al.*, 1997) they appear not to be used widely, probably because such programs are generally developed in house, and because the final image quality of some of these techniques is rather poor in comparison to that obtained using more traditional methods, probably in part because of a lack of further software development.



(a) (b) (c)

Figure 2: Examples of the three main approaches to facial approximation on the same skull **(a)** two-dimensional drawing; **(b)** three-dimensional clay sculpting; **(c)** computer generated (in this case also in two-dimensions).

Within each approach, three techniques of facial approximation may be used: (i) American method; (ii) Russian method; and (iii) combination method (Prag and Neave, 1997; Stephan and Henneberg, 2001; Taylor, 2001a). As previously mentioned, the American method employs average soft tissue thickness measures (see e.g., Aulsebrook *et al.*, 1996; Dumont, 1986; El-Mehallawi and Soliman, 2001; Garlie and Saunders, 1999; Helmer, 1984; Hodson *et al.*, 1985; Manhein *et al.*, 2000; Rhine and Campbell, 1980; Simpson and Henneberg, 2002; Smith and Buschang, 2001; Wilkinson, 2002) placed at various anatomical locations on the skull/face (Prag and Neave, 1997). The Russian method requires the build-up of facial anatomy from the skull primarily including the muscles of mastication and facial expression (Prag and Neave, 1997). The combination method is a mixture of both the Russian and American techniques (Stephan and Henneberg, 2001; Taylor, 2001a). See Figure 3 for examples of methods.



(a) (b) (c)

Figure 3: Examples of the main facial approximation techniques
(a) Russian technique, picture reproduced from (Gerasimov, 1971); **(b)** American technique, picture reproduced from Taylor (2001a) with permission by M. Micheletti; **(c)** combination technique.

In essence, the Russian technique relies on developing individual soft tissues from the skull while the American method relies on averages. Theoretically therefore, more emphasis is placed on "reading" skull morphology when using the Russian method than the American method. However, since few hard- to soft tissue relationships are known at present the soft tissue build up of the Russian method, including the build up of facial musculature, is largely subjective in reality (Montagu, 1947; Stewart, 1979b; Vanezis *et al.*, 2000).

There are only four muscles, the two masseters and the two temporalis muscles, which have both attachments to bone and contribute significantly to external facial appearance and bulk. All other muscles of the face contributing to outer face shape (primarily muscles of facial expression) have only one attachment to the skull and some, like Risorious and Obicularis Oris, have none and therefore cannot be estimated from the hard tissue alone. Additionally most of the muscles of facial expression are very delicate and leave little trace of their attachment and hence size and shape on the skull.

The American method bypasses much subjective speculation because standard average soft tissue depths are used, however, subjective interpretation is still required for the determination of face contours between known average soft tissue points. Consequently both the Russian and American methods involve a high degree of subjective interpretation.

The combination method seems optimal because it draws on the advantages of both the American and Russian techniques compensating to some degree for each techniques weaknesses (Prag and Neave, 1997; Taylor, 2001a). For example, the Russian method is limited because many facial soft tissues leave little clue as to their physical structure on the skull (Montagu, 1947; Suk, 1935) but the American method "makes up" for some of this by using average depths, which require little knowledge of the underlying tissues to represent points of facial contours at various anatomical locations. Similarly the Russian method compensates for weaknesses in the American method, like building "lip tension" into the mouth depending on occlusion type, which is not evident from average soft tissue depth measures alone.

While the combination method of facial approximation is an improvement on the Russian and American methods it continues to be severely limited because the amalgamation of the sub-methods does not fully compensate for the weakness they share: subjective interpretation. If facial approximation is to be an exact process, the Russian technique must predominate so exact face shapes are built. However, this requires comprehensive knowledge of how skull morphologies relate to soft tissue composition and shapes, which is largely unknown at this stage (see below). While few soft- to hard-tissue relationships are known it seems logical to use average methods in an attempt to restrict subjectivity. This approach (synonomous with the American technique) appears to be severely limited, however, because few people, if any, are average in all characters and, therefore, the use of such average methods may result in a face that is not representative of the actual target individual (person to whom the skull belongs) and may not be recognized (Brues, 1958; Stadmuller, 1922). An advantage of the average approach is that average faces are not distinctive (Bruce and Young, 1998) and therefore may look like many individuals possibly increasing the chance of non-distinctive target individuals being identified, but seems likely to create more leads for police to follow up. For more distinct skulls the average techniques seem unfavorable since the distinctiveness of the face is reduced making it harder to recognize as the target individual (Bruce and Young, 1998).

Exact methods for determining individual facial soft tissue from the skull, and hence pure Russian techniques, although optimal, may never be possible due to the complexity of soft- to hard tissue relationships as a result of chaos (small ordered changes that are dependent upon specific environmental conditions). Consequently, scientific generalizations based on averages are probably the most practical way forward because this extreme complexity is simplified, but are limited because they are only approximations of the truth and hence cannot be applied absolutely to individual scenarios. This approximation is often expressed by error ranges in scientific calculations that reflect the amount of chaotic variation, which is sometimes referred to as "random variation" even though this variation is not random nor caused by chance. Despite the limitations of scientific approaches (i.e., approximation of the truth (Popper, 1934; Popper, 1959; Popper, 1972)), systematic techniques are considerably advantaged in comparison to subjective methods because they are based on evidence, have known accuracies, and are repeatable within known ranges and can therefore be validated irrespective of the errors they may display. Consequently, it seems that combination methods of facial approximation, which include empirically based generalizations, while not optimal are the only way forward. As a result, facial approximation practitioners are likely to be forever forced into a tradeoff between "averageness" and the specific recognition of a unique individual. It is worth noting here that there have been attempts to individualize average soft tissue depths by basing them on skull sizes (Simpson and Henneberg, 2002; Sutton, 1969), but while this is an improvement of American techniques the measures remain limited because are still based on general trends which have error for individual prediction.

Controversy Surrounding Forensic Facial Approximation

Forensic facial approximation appears to be no easy feat. That is, the complete construction of a face that is representative of the individual to whom the skull belonged from the skull alone is an enormously zealous task. It is, therefore, not surprising that most of the criticism of the facial approximation method, and hence controversy, seems to be based on the fact that few specific hard- to soft tissue relationships are known and that much of the facial approximation process, therefore, is subjective and untested (Haglund and Reay, 1991; Montagu, 1947; Suk, 1935; von Eggeling, 1913). Consequently, the accuracies of the many subjective soft tissue predictions made in facial approximation are unknown and may be large.

Apart from the general size of the face and vague feature locations as indicated by the skull (e.g., general location of eyes "in" the orbits, nose about the nasal aperture and mouth over the teeth) the systematically tested and/or determined guidelines that exist for facial approximation include: two guidelines used to determine the width of the nose (Hoffman *et al.*, 1991); guidelines used to determine average soft tissue measures (Aulsebrook *et al.*, 1996; Dumont, 1986; El-Mehallawi and Soliman, 2001; Garlie and Saunders, 1999; Helmer, 1984; Hodson *et al.*, 1985; Manhein *et al.*, 2000; Rhine and Campbell, 1980; Simpson and Henneberg, 2002; Smith and Buschang, 2001; Wilkinson, 2002); some general information to determine ear height and angle (Farkas *et al.*, 1987); gross anatomy of the nose (Macho, 1986; Macho, 1989; Schultz, 1918); palpebral ligament attachment (Stewart, 1983); and muscle insertion at the mouth angle (Greyling and Meiring, 1992). Almost all other aspects of the face are subjectively determined using personal improvisation or previously suggested but *untested* soft tissue prediction guidelines, which most facial approximation practitioners appear to have followed blindly (since no tests have been published yet the guidelines are still used). It seems that one cannot expect a face, which is representative of the individual in question to be reliably built from a skull using the few scientifically tested guidelines mentioned above. Therefore, the scientific rigor of the facial approximation method that is often implied (Phillips and Snuts, 1996; Roy, 1992; Safe, 1991; Taylor, 2001a; Wilkinson, 2002) seems to be overemphasized (Haglund and Reay, 1991), and while facial approximation has been commonly referred to as a blend of science and art (Nelson and Michael, 1998; Phillips and Snuts, 1996; Roy, 1992; Safe, 1991; Taylor, 2001a; Wilkinson, 2002; Wilkinson and Whittaker, 2002), and even as "the scientific interpretation of the skull" (Wilkinson *et al.*, 2002) p.111, it is probably better described as much (much) more art than science.

It has even been found that some of the suggested subjective guidelines do not approximate the truth well. For example, the guideline that the height of the ear is equal to the height of the nose (Fedosyutkin and Nainys, 1993; Gatliff, 1984; Krogman, 1962) has been found to be infrequently correct since 95% of people have ears bigger than their noses (Farkas *et al.*, 1987), yet surprisingly, the guideline is still being recommended as a "rule of thumb" (Gatliff and Taylor, 2001). Additionally, the fact that many different soft tissue guidelines exist for predicting the same facial feature (Table 1) suggests that neither the soft to hard-tissue relationships, nor the reliabilities of the techniques are known (Stephan and Henneberg, 2001). It is also unlikely that these subjectively determined guidelines reliably account for individual variation because they have not been based on any empirical evidence. Consequently, it would not be surprising if most subjective soft tissue prediction guidelines were inaccurate, causing rare true positive recognition of facial approximations, and low recognition rates when they are identified correctly.

Table 1: Examples of multiple soft tissue prediction guidelines for single features

FEATURE	SPECIFIC SOFT TISSUE PREDICTION GUIDLINES
Mouth Width	1. Equal to the distance between the junction of the maxillary canine and the first premolar on each side (Gatliff, 1984; Krogman, 1962) 2. Equal to the distance between two perpendiculars dropped from the pupil centers (Caldwell, 1986; Krogman, 1962) 3. Corresponds to the distance between the mandibular second molars (Fedosyutkin and Nainys, 1993)
Mouth closure line	1. Anywhere in the region of the upper central incisors (Gerasimov, 1971) 2. Lower third of central maxillary incisors for females or lower quarter for males (George, 1987) 3. Equal to the line formed by the teeth when the mouth is closed (Fedosyutkin and Nainys, 1993)
Nose Width	1. Equals the width of the nasal aperture +10mm for whites and +16mm for blacks (Gatliff, 1984) 2. Equals the width of the nasal aperture +10mm for whites and +15mm for blacks (Schultz, 1918) 3. In Caucasoids the nasal aperture is approximately 3/5 of the total nose width (Gerasimov, 1971; Krogman, 1962; Prag and Neave, 1997)
Nose Projection	1. Is 3x the length of the nasal spine (Gatliff, 1984; Krogman, 1962) 2. Two tangents are projected, one following the last (distal) third of the nasal bones and the other following the general direction of the nasal spine. Where the tangents intersect indicates the tip of the nose (Fedosyutkin and Nainys, 1993; Gerasimov, 1971; Prag and Neave, 1997) 3. Is equal to the reflected profile line of the nasal aperture (Prokopec and Ubelaker, 2002)

It is also worth noting that much of the medical surgical literature on soft- to hard-tissue relationships has little application to facial approximation techniques because these studies determine what effect certain hard-tissue structures have on the soft tissue by altering the bone, often in conjunction with the treatment of "abnormalities" (e.g., Battagel, 1990; Hackney *et al.*, 1988; Kajikawa, 1979; Lin and Kerr, 1998; Ngan *et al.*, 1996; Wittbjer and Rune, 1989). The results of such studies, while useful to some degree, are limited when directly applied to the understanding of normal soft to hard-tissue relationships because the rules that apply when manipulating the soft/hard-tissue complexes (to achieve a desired result) may be different to those that operate between natural or unaltered soft and hard-tissues which are the focus of facial approximation.

Those who express some doubt about the abilities of the facial approximation method to achieve specific and purposeful facial recognition based on the evidence above have at times strongly criticized the method. For example, paper titles have included "Fallacies of anthropological identifications and reconstructions" (Suk, 1935), and "A study of man embracing error" (Montagu, 1947). Furthermore, three-dimensional clay reconstructions have been described as little more than "detective fiction" (Brues, 1958). At the other end of the spectrum, some seem convinced that facial approximation techniques do achieve their aim, for example, Prag and Neave (1997, p. 10) state: "the fact that the majority of the forensic reconstructions are recognized and identified demonstrates beyond a doubt that the technique works".

Several lines of evidence exist for assessing the ability of facial approximation to achieve specific and purposeful recognition: (i) forensic casework success; (ii) reported practitioner success; (iii) comparisons of facial approximations directly to target individuals; and (iv) recognition tests of facial approximations. This evidence is presented and evaluated below.

Evidence Indicating If Facial Approximation Achieves Its Goal, or Not

Forensic casework

Facial approximation has been successful in generating leads/tentative identifications in at least some forensic cases involving unidentified human remains. Published examples are: Cherry and Angel (1977); Farrar (1977); Gatliff and Snow (1979); Perper *et al.* (1988); Phillips *et al.* (1996); Prag and Neave (1997); Rathbun (1984); Stoney and Koelmeyer (1999); Suzuki (1973). This indicates that facial approximation does work, at least occasionally. However, the success of some individual facial approximation cases does not seem indicative of facial approximation accuracy because successes may be due to factors independent of the facial approximation, for example, contextual information (e.g., clothes and rings found at the crime scene) (Haglund and Reay, 1991) or chance (Stephan and Henneberg, 2001) and may be influenced by other factors, like broadness of media coverage (Haglund, 1998). Additionally, there seem to be many cases where facial approximation was not successful, e.g., Haglund et al. (1991). Furthermore, it also seems likely that many unsuccessful cases go unreported and/or unpublished, resulting in a biased account of facial approximation success when examining published case reports (Stephan and Henneberg, 2001).

Practitioner success

Some facial approximation practitioners claim high success rates indicating that facial approximation works, Gerasimov (1971) claims 100% success, Bender (Rubin, 1998) 85%, Wilkinson (Wilkinson and Whittaker, 2002) 75%, Gatliff (Gatliff and Snow, 1979) 70% and others claim more conservative but still high rates, for example Neave (Prag and Neave, 1997) 50-60%.

Reported practitioner success may be unreliable since the reporters (facial approximation practitioners) have a conflict of interest. They depend on success for attracting work and money. Therefore, they have to proclaim a high success rate, otherwise who would employ them? Some have reported that recognition rates have appeared to increase from around 30% in the late 1980's as competition between practitioners increased and peers started quoting their own (often higher) success rates (Taylor R, 2001 personal communication). Such trends are even evident within the last few years, for example reported success of R. Neave's method which is often described as either the "British" or "Manchester" method has increased from ~55% (Prag and Neave, 1997) to 75% (Wilkinson and Whittaker, 2002) over the last five years, apparently without any dramatic changes in the method since they have not been reported in the literature. Wilkinson and Whittaker (2002) even cites Prag and Neave (1997) in support for the claimed 75% success rate eventhough Prag and Neave (1997) state on p.33 that, "...the average success rate in most countries is between fifty and sixty percent, regardless of the techniques employed: in Mancehester the figure is no different".

Furthermore, it seems some facial approximation practitioners have taken a biased approach to reporting success rates selectively quoting lower success to methods the do not use and higher success for those they do. For example, although Gatliff has claimed a 70 pecent recognition rate (Gatliff and Snow, 1979), which Prag and Neave also acknowedge 72% (Prag and Neave, 1997), Wilkinson who uses the Neave method indicates that American method as practiced by Gatliff only achieves a 65% success rate, while the British method as used by Neave claims a 75% success rate (Wilkinson and Whittaker, 2002), even though Neave himself reports that the success rate of the "British" (or "Manchester") method is between 50 and 60% (Prag and Neave, 1997). Given the variability and inconsistency in reports of facial approximation success together with the fact that actual case numbers have not been quoted in support for the claimed success rates, it seems logical to suspect that the reported success rates have been summized by practitioners rather than precisely calculated from actual cases. Irrespectively, it seems that between well-known practitioners the success rates of facial approximation methods in general may have established an equilibrium at about the 50-70% success level, at least at the present time.

Direct comparisons

Comparisons between the facial approximation and the target individuals that have been identified are mostly encouraging, excepting earlier studies by Von Eggling (1913) and Stademuller (Stadmuller, 1922; Studmuller 1925). Many similarities have been reported to exist between facial approximations and their respective target individuals from subjective assessments (Gerasimov, 1971; Helmer *et al.*, 1993; Krogman, 1946; Prag and Neave, 1997; Suzuki, 1973; Taylor, 2001a). For example, Krogman (1946), p.17 reports a facial approximation that was "recognizable as that of the subject chosen". Suzuki (1973), p. 78, reports that "the resemblance between the two [a target individual and a facial approximation] was quite striking". Helmer *et al.* (1993), p. 236, conclude that "in general it can be said that at least a slight and often even a close resemblance was achieved" from the facial approximations to target individuals. Prag and Neave (1997), p. 35, 20 respectively, also report a "reconstructed face [that] bore an uncanny resemblance to the photograph [of the target individual]"; and of different approximations: "...the similarities between the faces and reconstructions were obvious to see and could not have been reached by chance". Vanezis and colleagues (1989) p. 70, also state that while "[facial approximation] can never produce a 100% accurate portrait...it will in the vast majority of cases, produce a head and face very similar to the original".

Although, as indicated above, recent trends for comparisons of facial approximations to target individuals is that they are similar (e.g., Helmer *et al.*, 1993; Krogman, 1946; Prag and Neave, 1997; Suzuki, 1973; Taylor, 2001a), older evidence suggests that they are not (Diedrich, 1926; Stadmuller, 1922; Studmuller, 1925; von Eggeling, 1913). For example, in von Eggeling's study two practitioners each produced a facial approximation of the same target individual but the facial approximations did not resemble each other or the target individual very much (von Eggeling, 1913). Stadmuller (1922) following facial approximation methods of Kollman and Buchly (1898) found facial approximations of two skulls to have little resemblance to their respective target individuals. In another study, Stadmuller (1925) found five facial approximations to bear little resemblance to their respective target individuals.

Additionally, reports of high resemblance of facial approximations to target individuals may be biased since in many publications the original facial approximations are retouched after a true positive identification has been made, for example facial approximations are given hairstyles and expressions that closely resemble their target individual (e.g., Gerasimov (1971) – see figures between pages 16 and 17). Helmer and colleagues (1993) even state that the addition of hair to facial approximations is "purely intuitive" and any resulting similarity is "only by chance", yet surprisingly the facial approximations illustrated by Helmer et al.

have hair types and styles remarkably similar to their respective target individuals (Stephan and Henneberg, 2001), which may account for their favorable results (see above).

Whatever the results of direct comparisons, the method appears to be limited irrespectively because facial approximation success depends on recognition, not similarity (Stewart, 1979b). Similarity, as measured by direct comparisons, often in the form of resemblance ratings, does not necessarily determine the "recognizability" of a face since similarity to non-target individuals is not accounted for and dissimilar faces may still be recognized correctly. The latter is evidenced not only in the recognition or pixilated images of faces, and caricatures (Benson and Perrett, 1991; Rhodes *et al.*, 1987), but also in the published literature where there are examples of facial approximations that look rather dissimilar to the target individuals, but which are recognized correctly (see e.g., Rathbun, 1984). As a result, direct comparisons may be of little use in assessing facial approximation accuracy and/or success (Stewart, 1979b).

Systematic tests

Some facial approximations have been correctly recognized above chance rates in systematic evaluations of the method that use face pools to test for recognition (Snow *et al.*, 1970; Stephan and Henneberg, 2001; van Rensburg, 1993). Snow et al. (1970) found two of two facial approximations to be identified above chance rates, which is evidence that facial approximation works and is accurate, however, recent empirical tests following strict facial approximation directions in the published literature have rarely resulted in above chance true positive recognitions of target individuals (Stephan and Henneberg, 2001). Even when facial approximations are recognized above chance in such studies, the rates are generally low - generally much less than 54% above chance, and being on average about 25% (Snow *et al.*, 1970; Stephan and Henneberg, 2001; van Rensburg, 1993). These recognition rates seem to be well below the 70% recognition rate reported in studies where the same individual is identified from images taken at different times even if view point and expression are constant (Hancock *et al.*, 2000).

Specifically, Stephan and Henneberg (2001) found only one of 16 facial approximations to be identified at a statistically significant rate (25%) above chance ($p<0.05$), with many non-target individuals identified for all facial approximations, some significantly above chance rates ($p<0.05$). Snow *et al.* (1970) found two of two facial approximations to be identified significantly above chance ($p<0.05$), one identified 12% above chance rates, the other 54% above chance rates. Again, non-target individuals were identified in both cases. In 'case 3', two non-target individuals (photos 4,6) were selected at rates close to that of the target individual (1) and appear to be above chance at the statistically significant level of $p<0.07$. Van Rensburg (1993) found that 15 facial approximations were, on average, identified at a rate 19% above chance ($p<0.05$), with the remaining identifications being of non-target individuals.

Therefore, while the face pool evidence may appear to be somewhat equivocal, there does seem to be a valid concern that facial approximation techniques may be inaccurate and unreliable, and may not achieve their goal of specific and purposeful facial recognition of target individuals, as indicated by others (Brues, 1958; Diedrich, 1926; Haglund and Reay, 1991; Montagu, 1947; Stephan and Henneberg, 2001; Suk, 1935). This concern is also strengthened by lack of "success" using direct comparison methods (Stadmuller, 1922; Stadmuller, 1925; von Eggeling, 1913).

A valid criticism of systematic tests of traditional drawing or sculpting methods that find negative results is that people constructing the faces from the skull have inadequate dexterity skills. It seems worth noting here that faces constructed by strictly following suggested soft tissue prediction guidelines while attempts are made by the practitioner to limit "open-ended" subjective interpretation (e.g., fine sculpturing of face contours like nasolabial folds and other wrinkles which cannot currently be determined from the skull) are likely to appear of lower "artistic quality" and less realistic than finely crafted faces because the latter include more precisely built features as a result of more subjective interpretation. Both types of facial approximations (high artistic quality and low artistic quality) may therefore be built by practitioners with similar dexterity levels, however, if less subjective interpretation is used when building a face from a skull using current methods, the "artistic quality" is likely to be less. There are, however, examples where "unlifelike" facial approximations are constructed even when much subjective interpretation is apparently used (e.g., the addition of wrinkles and other lines and grooves of the face), indicating a lack of practitioner dexterity (see e.g., Rathbun, 1984).

While criticism of systematic tests finding results of no difference (i.e., recognition rates not different from chance) on the grounds of limited dexterity may be justifiable as mentioned above, dismissals of negative results, as Gerasimov (1971) and Prag and Neave (1997) have done for von Eggeling's (1913) research, on these grounds is not appropriate. Firstly, many facial approximations made with limited dexterity and poor "artistic style" and/or dexterity are correctly recognized (see e.g., Rathbun, 1984), indicating that dexterity is not necessarily the main contributing factor to facial approximation success. Secondly, the inaccuracy of the method may be due to the use of inaccurate soft tissue prediction guidelines rather than a lack of dexterity and/or "artistic style", which perhaps would not be surprising given that most techniques used to estimate the soft tissues from the skull are subjective and untested, being based on little, if any, empirical evidence (Brues, 1958; Montagu, 1959; Stephan and Henneberg, 2001; Suk, 1935). Thirdly, because many facial soft tissues cannot yet be determined from the skull and can only be represented subjectively, it seems unlikely that *anyone* (even those with good dexterity and artistic sense) can create facial approximations from the skull alone, that can be reliably, specifically, and purposefully recognized correctly.

Despite the apparently extreme limitations of the facial approximation method even when a complete skull is present to work from (see above) some practitioners have taken the facial approximation process a step further by working from crania alone (skulls without mandibles) (e.g., Gatliff and Taylor, 2001), or even without a skull (!), using cranial measurements (e.g., Jackson, 1996). The logic here is inescapable, for without the bone structure itself these facial approximations (or parts thereof) must be highly, if not purely, subjective. It therefore seems extremely unlikely that purposeful and specific recognitions could be reliably made from such facial approximations. Successes in such scenarios seem logically likely to arise from factors independent of the facial approximation, like chance (Stephan and Henneberg, 2001) or contextual information (Haglund and Reay, 1991).

It seems strange that while Gerasimov (1971) and Prag and Neave (1997) refute von Eggeling's (1913) conclusions that facial approximation is an inaccurate technique, they do not provide detailed descriptions of their methods or justifications for them in their major books on the topic. For example, Gerasimov in "The Face Finder" (1971), pp. xxi-xxii, describes the technique briefly in 10 pages (5%) of a 199 page book dedicated to the topic and states that "The reader will not find in this book any detailed discussion about the justification for the work of reconstructing the appearance of Man from his bones. Not words but concrete examples will show that reconstructions of recent Man have proved the reliability of the method in crime detection." However, as we have seen above case examples provide limited evidence that the methods work since the cases presented may be selective and success may be achieved independently of the facial approximation. Prag and Neave (1997) also appear to briefly gloss over the techniques of facial approximation in "Making Faces", dedicating less than 10 pages (4%) of the 256 page book to specific methods. The dedication of such small segments of writing (~10 pages) to techniques that take on a huge task – rebuilding a face from the skull alone - seems less than sufficient. Gerasimov (1971), p. 54, states "...we offer some practical information about the reconstruction of individual face features. Of course, we can only touch on certain points and do not claim to exhaust the subject". But, one has to wonder why "of course" he can only touch on "certain points" after all the topic of the text is facial approximation and he has an entire book to describe it if he wants. One is certainly left wondering if the techniques the authors mention are the only ones they use and if recognizable faces can really be built from skulls using so few soft tissue prediction guides?

Taylor (2001a) provides a more detailed description of facial approximation methods in several chapters in her book, but the guidelines she uses are those common in the literature, which again are relatively few, and there is no attempt to demonstrate their empirical validity. Other articles describing facial approximation techniques (e.g., Caldwell, 1986; Krogman, 1962; Stewart, 1979b; Taylor and Angel, 1998) also frequently cross-reference common information between texts without providing much new detail or empirical justifications. Consequently, it does seem that few other soft tissue prediction guidelines exist for facial approximation in addition to those originally cited by Gerasimov (1971) and Prag and Neave (1997). If there were other guidelines that did exist but have been unreported as Gerasimov (1971) suggests, it seems they have been lost since they have not resurfaced in the facial approximation literature over the last 40 years – or maybe they did not even exist initially. The use of these rather few published guidelines for determining an entire face and the fact that most of these guidelines are subjectively determined and empirically untested does not encourage much trust in the methods of facial approximation. Reports that facial approximation techniques are inaccurate (Brues, 1958; Diedrich, 1926; Haglund and Reay, 1991; Montagu, 1947; Stadmuller, 1922; Stadmuller, 1925; Stephan and Henneberg, 2001; Suk, 1935; von Eggeling, 1913) also seem unsurprising in this context.

Since facial approximation is a potentially useful identification tool there seems to be a definite need for current methods to be tested, evaluated and possibly improved upon. This thesis sets out to evaluate two-dimensional facial approximation methods, however, many of the tests conducted have relevance to all forms of facial approximation, wether two-dimensional, three-dimensional, sculpted in clay or computer generated. The research section of this thesis begins with a study on photogrammetric methods, since these methods form the basis of other studies in later chapters. The photogrammetric method chapter is followed by tests of mouth width prediction guidelines, the superciliare prediction guideline, pronasale and nose projection guidelines, and the traditional eyeball projection guideline. Next, a study on the systematic generation of average face anatomy is presented along with a study indicating the sample sizes needed for creating reliable average faces. The thesis then presents research that determines the best way to assess facial approximation accuracy and uses this information in the next chapter to determine what the ceiling recognition rates of 2D facial approximation are by warping average face anatomy to exact face shapes. The final research chapter takes all the information obtained in previous chapters and puts it together, determining the "recognizability" of two-dimensional facial approximations created using "objective" methods (i.e., average faces and empirically derived soft tissue prediction guidelines which reduce subjectivity). This chapter also compares these recognition rates to "recognizabiltiy" levels obtained using traditional subjective and empirically untested methods and those from tests of ceiling recogntion. The thesis then concludes with a general discussion on the results found in the previous chapters and presents a summary on facial approximation accuracy, its usefulness in forensic science, consequences for professional and public perceptions of the method, and future research directions that appear to be pressing.

This thesis includes 10 "independent" research studies, 5 of which have already been published as 6 full-length papers in peer-reviewed forensic science journals, and another is currently in review. Reprints of full-length paper publications are presented in the appendices of the thesis.

Prediction of Soft Tissue Anatomy

Standardized Craniofacial Photography

This chapter forms the basis of a technical note that has passed initial peer reviews in *Plastic and Reconstructive Surgery* (see Appendix 1).

Introduction

Photography is principally a useful technique because it provides a visual record of the physical environment. Furthermore, it is possible to take measurements from photographs (photogrammetry), which is advantageous in the study of living human subjects because: (i) measures can be taken without any risk of soft tissue depression and hence inaccurate measurement; (ii) measurements can be made on completely static subjects; (iii) measurements can be precisely taken (with sub-millimeter accuracy either electronically or using calipers) and are primarily only limited to the resolution of the image removing the need for somewhat more cumbersome instruments such as spreading calipers; and (iv) the records are semi-permanent to permanent, enabling additional comparisons, or measurements, to be made at a latter time in exactly repeatable scenarios. As a result, photographic methods have been established as being an indispensable tool for recording human anatomy in medical and dental surgery (especially pre- and post- operatively); anthropology; and psychology. It is not surprising, therefore, that many papers have been written on various photographic methods and rigs (DiBernardo, 1991; DiBernardo *et al.*, 1998; Disaia *et al.*, 1998; Edgerton *et al.*, 1970; Farkas *et al.*, 1980; Ferrario *et al.*, 1995; Fricker, 1982; Fricker, 1985; Galdino *et al.*, 2001; Kesselring, 1985; Morello *et al.*, 1977; Nechala *et al.*, 1999; Ras *et al.*, 1995; Ras *et al.*, 1996; Thomann and Rivett, 1982; Thomas *et al.*, 1980; Zarem, 1983a; Zarem, 1983b).

For photographic comparisons to be valid, photographic conditions like lighting, pose, subject-camera distance, lens focal length, film type, and film processing must be tightly controlled and consistently applied across photography sessions. Such repeatability may be obtained by using appropriate equipment and standardized techniques (Dobrostanski and Owen, 1998; Gavan *et al.*, 1952). Generally, the best accepted protocols for standardizing these variables in craniofacial photography include: fixed, overhead studio flash units, positioned either at the same distance from the subject or with one slightly closer to give fine highlights (Dobrostanski and Owen, 1998; Zarem, 1983a); subjects photographed in the natural head position when standing, since it is highly repeatable and displays individuals as they usually appear in life (Cooke and Wei, 1988b; Moorees and Kean, 1958; Moorees *et al.*, 1976); use of a large focal length of lens (90 to 105 mm) and subject camera distance of about 1m since this represents individuals as in normal social contact with the features well proportioned (Dobrostanski and Owen, 1998) (a 105mm lens appears to be the most commonly used focal length lens in conjunction with a subject-camera distance of 1-1.5m (DiBernardo *et al.*, 1998; Dobrostanski and Owen, 1998; Fricker, 1985; Morello *et al.*, 1977; Thomas *et al.*, 1980; Zarem, 1983a)); and the use of Ektachrome slide reversal film, since it can be processed in most labs (DiBernardo *et al.*, 1998; Galdino *et al.*, 2001) and requires only one stage of development in comparison to colour prints (Dobrostanski and Owen, 1998). It should be noted that Kodachrome is probably superior to Ektachrome since these films are more fade resistant (~50 years in comparison to ~10 years)(Thomas *et al.*, 1980), however, Kodachrome is limited since it can only be processed at specialized laboratories (DiBernardo *et al.*, 1998).

Despite the ability to overcome variables like lighting and pose with standardization, there are some limitations inherent to the photography method that must be recognized. These limitations also severely affect the accuracy of photogrammetric techniques, which is particularly relevant to craniofacial photography (Chapple and Stephenson, 1970; Dickason and Hanna, 1976; Farkas, 1994b; Farkas *et al.*, 1980). The limitations include distortions as a result of perspective (Gavan *et al.*, 1952), as well as the impossibility of locating on 2D images points that can only be located in three-dimensions (Farkas, 1994b).

Perspective distortion is problematic because it causes a false impression of the size of objects (always being smaller than in reality) since the camera cannot "see" the true edges of three-dimensional objects (Gavan *et al.*, 1952) (see Fig. 4). Since perspective distortion is minimized at infinity, small subject-camera distances will exaggerate the distortion (Gavan *et al.*, 1952). It is therefore, best to use large subject-camera distances, and large focal length lenses. Figure 5 illustrates differing face distortions as a result of different focal length lens at differing distances.

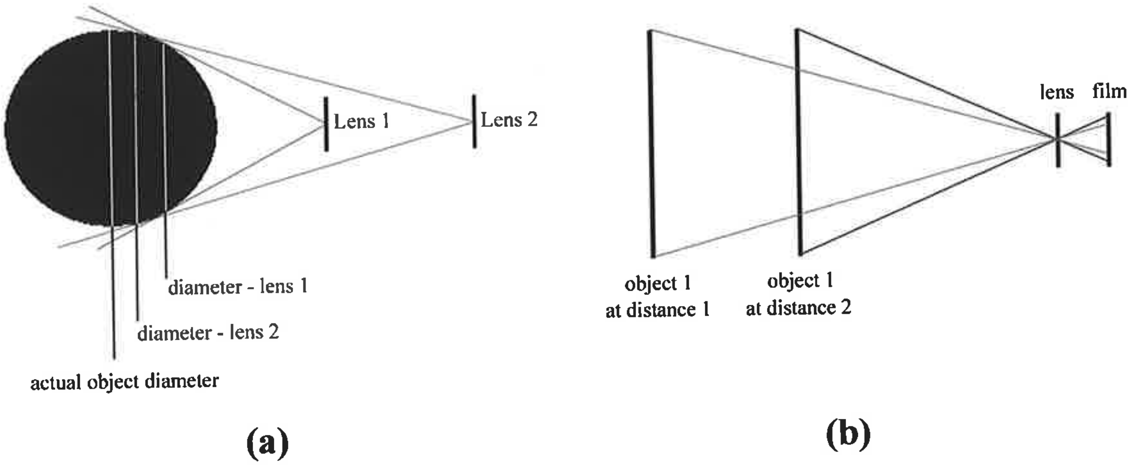


Figure 4: Distortions inherent in the photography method as a result of subject camera distance **(a)** incorrect edge representation of the same object; **(b)** image enlargement of the same object.

(a) (b) (c)

Figure 5: Frontal face images showing distortions as the result of different focal length lenses All photographs have been taken so magnification is consistent, i.e., closer subject-camera distances for smaller focal length lenses. **(a)** 28mm focal length lens; **(b)** 55mm focal length lens; **(c)** 105mm focal length lens, images reproduced from Dobrostanski and Owen (1998) with reprint permission by Hodder Arnold.

Magnification distortions are also a result of perspective effects, which cause objects closer to the viewpoint (or camera) to appear larger than objects further away (Fig. 4). Again, magnification distortions, like other perspective distortions are increased when using small subject-camera distances and small focal length lenses (Gavan *et al.*, 1952). It is worth noting here that the problem of magnification cannot be easily resolved by inclusion of a scale since any feature not in the same focal plane cannot be exactly measured. This can be overcome somewhat by taking both frontal and profile photographs that enable determination of feature distance from either camera. However, since some measurements must be taken across several planes it is difficult to adjust measurements precisely.

The final limitation is a result of the 2D nature of photographs. In craniofacial photography some traditional anthropometric landmarks may not be distinguishable because the landmarks maybe hidden by other features (Farkas, 1994b). For example, the midline may be obscured in profile by more lateral features that are located more anteriorly, e.g., if the eyebrows project in front of glabella. This may also happen in frontal views since landmarks may be obscured by more anterior and laterally located points, e.g., gonion may be hidden by the cheeks. In addition, some features, despite being visible, may not be possible to determine since the photograph is a two dimensional image (Farkas, 1994b). For example, the determination of pronasale (and most other mid line points) in the frontal view. Some landmarks can also only be determined by palpation (e.g., orbitale) and therefore they cannot be determined on photographs unless palpated and marked on the face beforehand (Farkas, 1994b).

Despite these limitations, craniofacial photography has the distinct advantage that actual three-dimensional facial anatomy can be visualized in two dimensions. This is useful since it allows standard anthropometric measures/values to be seen in a "biological" rather than "mathematical" context. Three dimensional laser scanning and computer graphic applications are advantaged in comparison to 2D photographic techniques since the above limitations can be overcome. However, 3D techniques are expensive and are not (yet) widely used, unlike 2D methods.

Two options exist for craniofacial photography at this time. There are traditional chemical methods that typically use a 35mm camera body and film, or there are digital techniques that use a digital camera and computer hardware and software. Conventional 35mm cameras have the advantage that image resolution is very high, techniques are widely practiced, and hard copy images can be digitized if needed. Limitations include: time delay in image development; and semi-permanence of hard copy images (10-50years depending on film type and storage). In comparison, digital techniques are advantageous since images are immediately available, can be of high resolution, and can be permanently stored without degradation. However, digital methods are limited for some established laboratories since upgrading is required from conventional techniques, which can be expensive.

Although many papers have been published on standardized photography techniques in medicine (e.g., DiBernardo, 1991; DiBernardo *et al.*, 1998; Disaia *et al.*, 1998; Edgerton *et al.*, 1970; Farkas *et al.*, 1980; Ferrario *et al.*, 1995; Fricker, 1982; Fricker, 1985; Galdino *et al.*, 2001; Kesselring, 1985; Morello *et al.*, 1977; Nechala *et al.*, 1999; Ras *et al.*, 1995; Ras *et al.*, 1996; Thomann and Rivett, 1982; Thomas *et al.*, 1980; Zarem, 1983a; Zarem, 1983b) it is surprising that none have reported on or displayed images that indicate the repeatability of sequential photographs taken when using the methods described, since repeatability is a main priority of standardized photography.

This study aims to determine the repeatability (intra-observer error) that can be expected when taking multiple photographs on a newly built photography rig at the School of Dental Science (the University of Melbourne) that uses conventional 35mm cameras. The rig has been designed specifically in an attempt to enable simultaneous, high quality photographs to be captured in both front and profile views of the face, in a realistic, fast, easy and accurate manner. The photography rig is a significant improvement on others since it uses as novel projected light range finding system that allows subjects to be positioned in the natural head position at constant subject-camera distances.

Methods

Description of the photography rig

The rig is a modified version of that already described by Dobrostanski and Owen (1998). The rig frame consists of a rigid stand that supports two cameras (Fig. 6). One camera is positioned to capture a frontal image of the subject, the other a right side profile. The stand is made of aluminum (square cross-section tubing) that is firmly secured to the floor. The stand is 2138 mm high and has two horizontal beams that are approximately 1650 mm in length. These beams are positioned at 90 degrees in relation to one another (producing an "L" shaped gantry) and each holds a motorized Nikon SLR camera that is vertically mounted (Fig. 6& 7). The beams are secured to the main frame, which can be moved to accommodate subjects as short as ~1400mm or as tall as ~2170mm. However, it is worth noting that shorter subjects can be appropriately and easily positioned with the help of a raised step/platform.

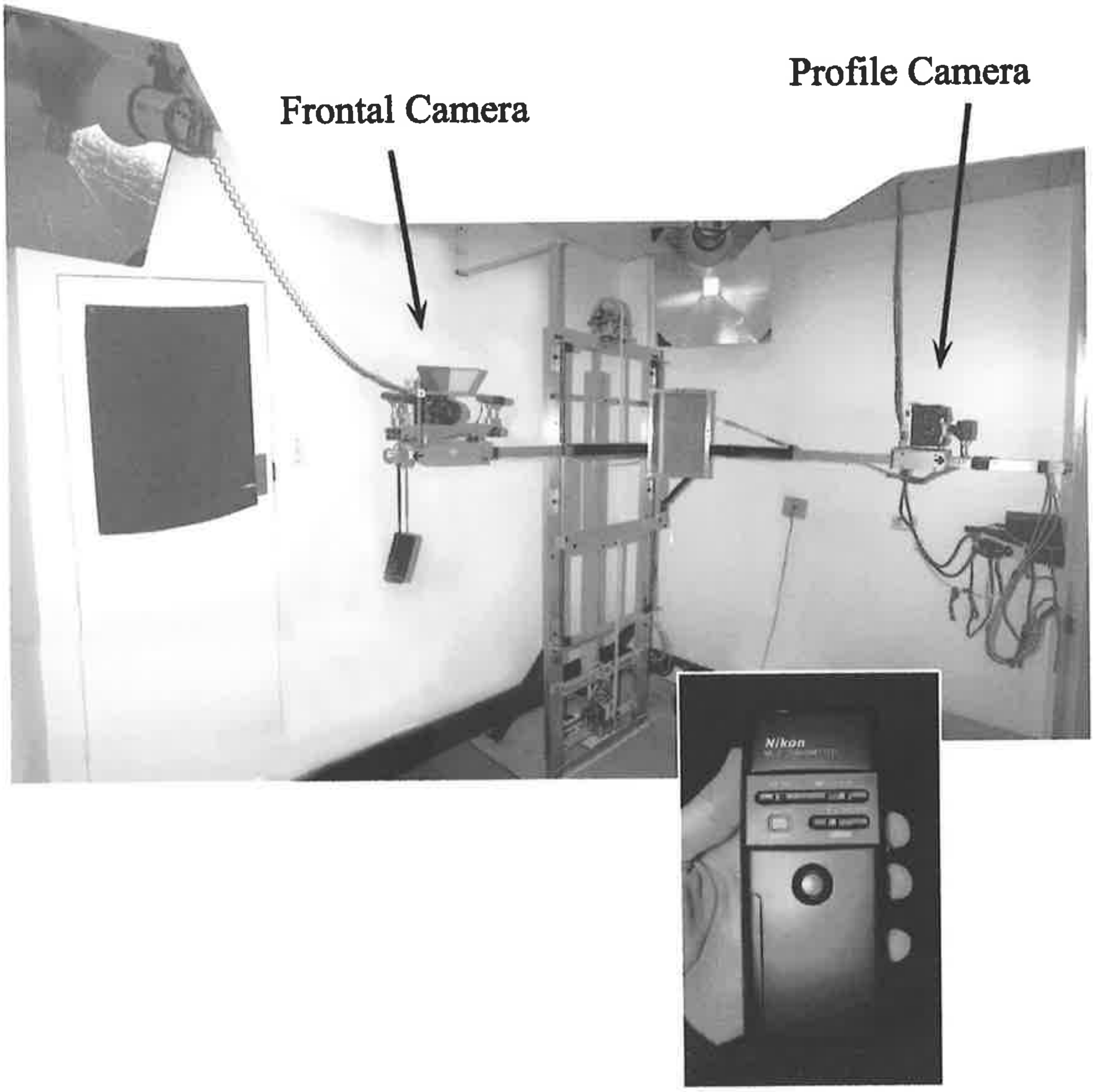


Figure 6: Photography rig.
Note that the picture is not to scale. Inset shows infrared remote.

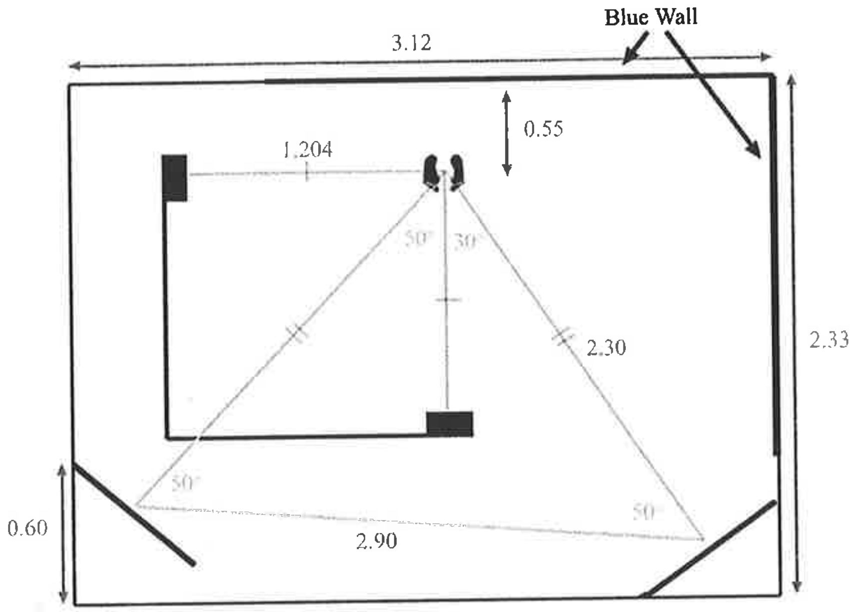


Figure 7: Floor plan of the photography rig
Units are meters.

To allow fine adjustment of the height of the cameras, the beams are counterbalanced by a lead weight. An automatic clutch prevents any vertical creep of the beam once the height has been set. The extreme vertical movement of the beams is limited by cushioning stoppers. Each horizontal beam is fitted with an adjustable horizontal sliding camera platform that enables precise positioning of the cameras.

The cameras used

The cameras consist of two Nikon FM-2 35mm bodies, each fitted with an E2 grid-focusing screen. Each camera body is coupled to an MD12 motor drive. The battery packs of the drives have been replaced with a 12.8 V, 2 A, DC external power supply connected with an appropriate length of cable. The cameras are fitted with Nikon 105 mm 1:2.8 Macro Nikor lenses. The faces of subjects are positioned in the camera viewfinder at a distance of 1204mm from the film plane in frontal view, and ~1204mm from the film plane in profile view (this distance is only approximate since the profile camera is fixed and therefore distances may vary slightly depending on subject head width), giving photographs that are representative of natural images seen during normal social contact.

The camera shutters are activated by an infrared triggering system. This system consists of a receiver that is mounted to the profile camera hot shoe. The receiver triggers the shutters via electrical contact on the motor drives. The cables that connect the two cameras to the infrared receiver are equal in length so that the camera shutters activate synchronously. The infrared transmitter is a remote, battery-powered, directional, hand held unit (inset Fig. 6).

The lighting used

Two Elinchrom Prolinca 2500 self-contained studio electronic flash units are used to illuminate the subject (Fig. 6 & 7). To avoid harsh shadows, foil reflectors are used to bounce the light from the flash units towards the subject. The flash units and reflectors are placed higher than the subject, with one light being slightly closer to the subject than the other. One flash unit is positioned 50 degrees to the right of the subject at a distance of approximately 2300mm. The other unit is placed 30 degrees to the left of the subject also at a distance of approximately 2300mm (Fig. 7). This arrangement is sufficient to permit the use of an f-stop of f/16 with Ektachrome E-200 slide reversal film. The slightly closer positioning of the right flash unit results in slight highlights on the subject giving a three-dimensional effect to the final images (Dobrostanski and Owen, 1998). The use of a focused distance of about 1.2m and an f-stop of f/16 gives a depth of field from about 1.15m to about 1.26m.

Alignment of the subject

A mirror is placed above the lens of the frontal camera and is angled anteriorly and inferiorly at 5 degrees from the vertical so that the subject can see the reflection of their eyes in the mirror and adapt a natural head position, which is reported to be reproducible with little variance (Cooke, 1988; Moorees and Kean, 1958). Two non-laser light pointers are used to centrally position the subject within the field of view of each camera. One pointer lamp is placed on each side of the frontal camera and is horizontally angulated so that the projected beams of light converge at the point of sharpest focus from the cameras (Fig. 8). Light emitted by the pointers is not potentially harmful laser light but incandescent light from fitted 3V Rowi 528 globes powered by an external 3V power supply. The left pointer (from the subject) projects a "<" shaped image, the right a ">" shaped image. The two range finding pointers are calibrated so that the "v" shaped images form an "x" at the point of sharpest focus. If the two "v" shaped images overlap the subject is too close to the lens and if the two "v" shaped images are spread apart the subject is too far away (Fig. 8). When the camera shutters are activated, the light emitted from the flash units overpowers the pointer lights so the "v's" are not registered on the film. The focus of the range finding pointer lamps can be adjusted to ensure the projected "v's" are crisp at the desired distance.


(a) (b) (c)

Figure 8: Pointer alignment of subject
(a) too close to camera; **(b)** correct position; **(c)** too far from camera. (To enable the light aligners to be seen photographs were taken without a flash using a camera positioned above the frontal mirror with a 20-80mm lens).

The profile camera is positioned so that when the subject is correctly aligned, its field of view will include all of the subject's profile (from behind the ear to in front of the nose). To assist positioning of the subject "two cardboard feet" are stuck to the floor to give an approximate indication of the focal planes of the cameras (Fig. 7). The wall in front of each camera is painted a medium sky blue with a non-glossy paint to give a uniform background that reduces glare and is easily differentiated from the subject in the final images (DiBernardo *et al.*, 1998; Dobrostanski and Owen, 1998; Galdino *et al.*, 2001; Thomas *et al.*, 1980).

Subjects are positioned following methods of Moorees and colleguaes (Dobrostanski and Owen, 1998; Zarem, 1983a). Subjects are asked to stand on the cardboard feet in an upright position and to look directly at the wall in front of them. At this point the "L" shaped gantry is lowered from its highest position to a position determined by the placement of the light pointers at glabella. The subject is then asked to look at their eyes in the mirror above the lens of the frontal camera. This helps to position their head in the natural head position if they were not already in it (Dobrostanski and Owen, 1998). The subject is then asked to gently move forwards or backwards until the pointers form an "x" on a reference point, e.g., glabella, the positioning of which is controlled by the operator of the rig by adjusting the vertical height of the camera beams. Glabella is used here as the reference point, however, sellion would also be appropriate. Positioning of the subject at glabella/sellion is, however, a compromise between easy and rapid operation of the rig and precise focusing, which should take place at the eyes (Dobrostanski and Owen, 1998). Once the subject is correctly positioned the cameras can be activated (simultaneously) from the remote unit without the need for the photographer to view the subject in the camera viewfinder.

In day-to-day use of the rig there is no need to adjust the pre-selected focus of the cameras/pointer aligners because each subject adjusts their distance with respect to the cameras using the range finding light pointers. All that has to be done when photographing multiple subjects is to position the first using the projected light range finding system, activate the shutters and the flash units using the infrared remote, and repeat the process with the next subject. Since the SLR cameras automatically wind on, subject throughput is very rapid.

Experiment 1

This experiment was set up to determine the scaling factor required to adjust photographs to life size. Although this could be done theoretically, here it was determined experimentally so actual functional values were obtained since these have the greatest significance (i.e., theoretical values may differ to functional results). A ruler was positioned by fixing it to a camera tripod with Bostik Blu Tack®. The ruler was photographed and then removed. The ruler and tripod were then repositioned, photographed once again, and removed. This procedure was repeated until seven sequential photographs were obtained on the same day. Rulers were photographed in horizontal and vertical positions, but not on the same day. Photographs were taken using 200 ISO Ektachrome slide reversal film.

Once the slides were processed they were mounted, and scanned into a computer using a Nikon® SF-2000 slide scanner. The resultant pictures were 1,200 pixels in width, 1,803 pixels in height (48.0 x 72.1mm) and were saved in TIFF format. Several preferences in Nikon® Scan were also selected to produce images that were considered to be highly representative of the original slide image (e.g., clean image function was set to normal, bit depth was set to 8, multisample was set to 4x (fine), interpolation was default, colour space was RBG Adobe® 1998, and autofocus and autoexposure were turned on), these preferences were held constant during scanning of the slides.

The scanned picture of the ruler was then viewed and measured in Adobe® Photoshop® 6.0. Measurements were made across 100mm sections of the ruler. This distance was chosen since it is relatively large, and is comparable to, or larger than, many measurements of the face such as those of the nose, mouth and eyes. Averaging the measurements, and dividing the true object size by this value (100mm) allowed the calculation of the average magnification factor needed to obtain actual size for images placed at the point/plane of sharpest focus.

Experiment 2

The aim of this experiment was to determine what the magnitude of changes in object magnification were, resulting from the positioning of objects about the plane of sharpest focus. The experiment consisted of photographing a ruler (mounted as in experiment 1) at differing positions around the point of sharpest focus. This was done by placing the ruler on a tripod, which had an adjustable platform and a stationary base. The platform was then moved at 10mm intervals, in various directions, and the ruler photographed. Measurements of the photographed ruler, across 100mm, made it possible to determine changes in magnification due to positioning about the focal plane.

Experiment 3

This experiment was conducted to determine the error in repeated photography of the same human subject on the same day. The subject was positioned as previously described and photographed with a neutral facial expression with his mouth closed, on six occasions. Superimpositions were made of the photographs to indicate repeatability. Also, numerous anatomical landmarks, as defined by Farkas (1994), were measured on the images and compared to determine the intra-observer error of photography. Coefficients of variation were generated by dividing the standard deviation by the average and multiplying by one hundred.

Experiment 4

This experiment was the same as experiment 3 except that sequential photographs were not taken on the same day. Photographs were taken, on average, every 5.2 ± 4.2 days.

Results

Table 2 shows the results of the unadjusted measures taken across 100mm of photographed rulers from the scanned slides. The average value of the 100mm as measured from the slides was 20.1 (mm), giving a magnification factor of 4.975124. The average value of 20.1 is an approximation of the true value since the standard deviation (0.45mm) indicates at a 99% confidence level that the true value could lie between 18.8 and 21.4mm. Therefore the maximal error in using the scaling value of 4.975124 as determined by the average of 20.1 could be up to 6.5mm, but is probably less.

Table 2: Measures made on repeated photographs of rulers (mm) to calculate scaling factor
All measurements are made across 100mm as determined by the rulers.

Trial	Frontal Camera Horizontal Ruler	Frontal Camera Vertical Ruler	Profile Camera Horizontal Ruler	Profile Camera Vertical Ruler	Average
1	19.70	20.70	20.00	19.80	20.05
2	19.60	20.70	20.10	19.80	20.05
3	19.70	20.70	20.30	19.70	20.10
4	19.70	20.70	20.40	19.70	20.13
5	19.50	20.70	20.40	19.80	20.10
6	19.70	20.70	20.40	19.80	20.15
7	19.70	20.70	20.70	19.80	20.23
Average	**19.66**	**20.70**	**20.33**	**19.77**	**20.11**
SD	**0.08**	**0.00**	**0.23**	**0.05**	**0.10**

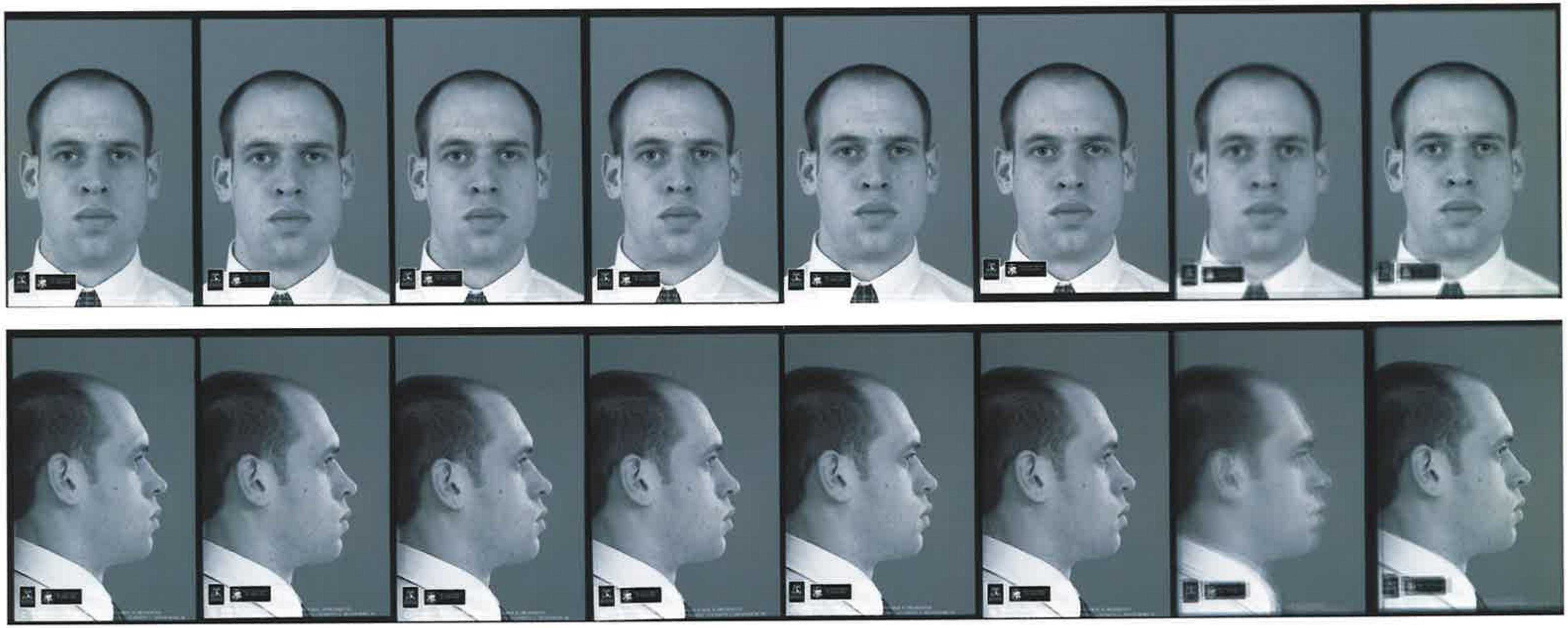
Table 3 shows that movement of the ruler by 10 mm in any direction generally resulted in a 1mm change of object magnification as measured across 100mm on the photographed ruler. This is equal to a 1% change in object size for every 10mm the object is closer to, or further from, the camera. As expected, the increase was positive when the movement was toward the camera and negative when away from the camera.

Table 3: Magnification effects from movement of the ruler about the point of sharpest focus
All measures (mm) are based across 100mm.

Ruler Position	Full Face	Profile
Forward 40mm	103.9	100.5
Forward 30mm	102.9	100.5
Forward 20mm	101.9	100.0
Forward 10mm	101.0	100.5
Point of sharpest focus	100.0	100.0
Back 10mm	99.0	100.0
Back 30mm	97.1	100.5
Back 40mm	96.1	100.5
Left 20mm	100.5	98.5
Right 20mm	100.0	102.0

Figures 9 and 10 show the six sequential facial photographs taken of the same subject. It can be seen from these images that repeatability is high although not exact. True image borders are indicated by the black borders at the top and the right of the images, while the left and bottom borders are artifactual (e.g., slide mount borders). The straight superimposition (Superimposition 1, Fig. 9 and 10) is based on true image borders and illustrates repeatability of sequential photographs. It can be seen from this image that variation in positioning was minimal, but present. It can also be seen from Figures 9 and 10 that most variation in subject positioning occurred in the anterio-posterior plane since repeatability was generally less for profile images. In Figure 10 the photograph taken eight days after the first, illustrates the largest deviation from the average position and may be due to the subject leaning forward in the instant between correct positioning and the pressing of the remote button to activate the camera shutters. Also displayed in Figures 9 and 10 are superimpositions after alignment on the eyes of the subject. From these superimpositions it can be seen that the slight deviations in subject positioning had minimal effects on face morphology, even when photographs were taken on different days (Fig. 10).

Measurements of the faces taken on same and different days are represented in Tables 4, 5, 6, and 7. These tables illustrate highly repeatable measures that appear to be precise for both conditions (same and different days) since coefficients of variation are low. The highest variation appeared to be in the natural head position, as measured by the angle formed between t-ex (tragion and exocanthion) and the horizontal plane in profile, (coefficients of variation of 6.3 for same day photographs and 6.7 for different day photographs) and in profile measurement of the stomion to labiale inferius (coefficients of variation of 3.9 for same day photographs and 7.5 for different day photographs). However, actual variations were low, for example, the maximum variation for the angle formed between the horizontal and the plane from tragion to exocanthion (t-ex) was 3.1 degrees (same day photographs), which is comparable to the repeatability of natural head position reported elsewhere (Cooke, 1990; Peng, 1999; Siersbaek-Nielsen and Solow, 1982; Solow and Tallgren, 1971).



Trial 1 Trial 2 Trial 3 Trial 4 Trial 5 Trial 6 superimposition 1 superimposition 2

Figure 9: Photographs of the same person taken on the same day
"Superimposition 1" shows a straight superimposition of the 6 photos. "Superimposition 2" shows the superimposition of the 6 photos aligned on the eyes. True image borders are indicated by the black top and right borders, while left and bottom borders are artifactual.

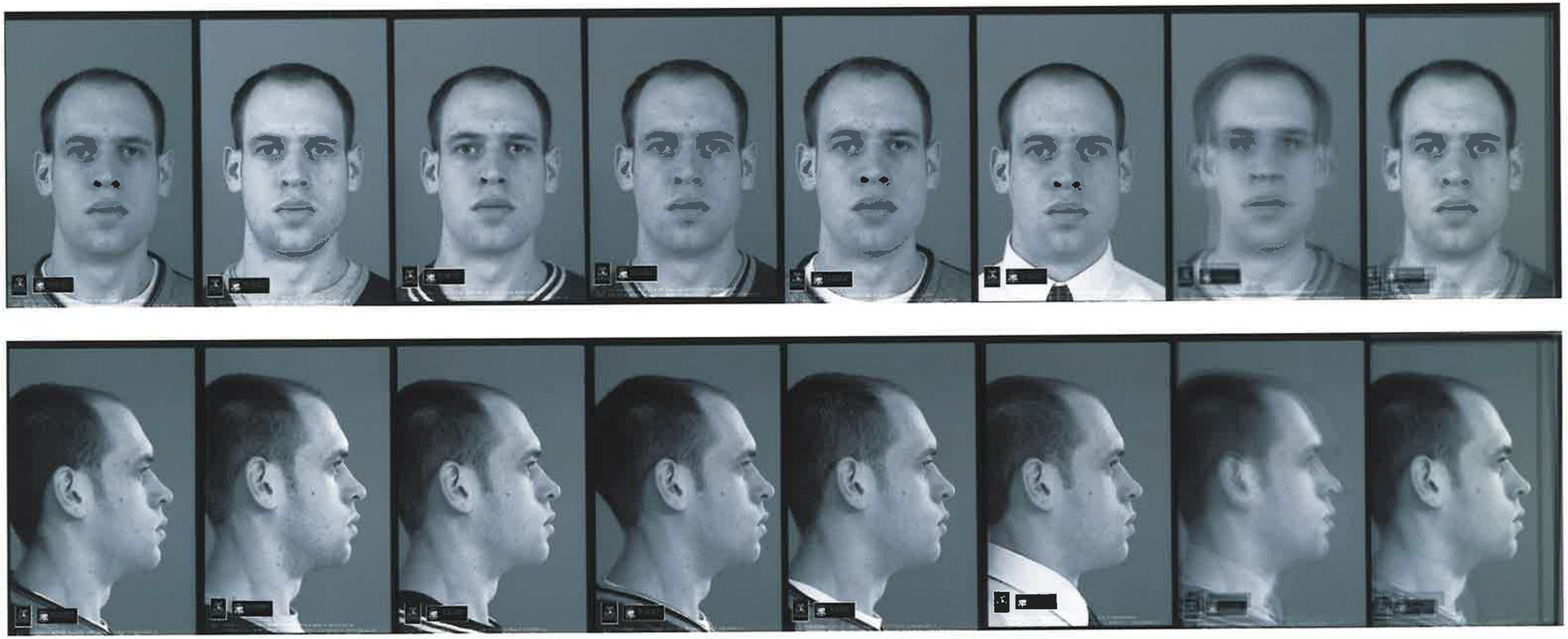


Figure 10: Photographs of the same person taken on different days
"Superimposition 1" shows a straight superimposition of the 6 photos. "Superimposition 2" shows the superimposition of the 6 photos aligned on the eyes.
True image borders are indicated by the black top and right borders while left and bottom borders are artifactual.

Table 4: Measurements (mm) of the subject's face, from same day frontal photographs.

Trial	ch-ch	cph-cph	al-al	p-p	en-ex [r]	en-ex [l]	en-en	ex-ex	sa-sa	sba-sba	sa-sba [r]	sa-sba [l]	ls-li	ls-sto	li-sto	sn-sto
1	49.8	16.9	41.3	64.2	26.9	26.9	32.8	86.1	162.2	149.3	59.2	60.7	22.9	10.4	12.4	28.4
2	49.3	15.9	40.3	64.7	27.4	26.9	32.8	87.1	163.7	149.8	58.7	60.7	23.9	10.9	12.9	28.9
3	49.8	17.4	39.8	64.7	27.4	26.9	32.8	86.6	162.7	149.3	58.7	60.2	23.9	10.9	13.4	27.9
4	48.8	16.4	39.8	63.7	27.4	26.4	32.3	85.6	161.7	148.3	58.2	60.2	24.9	10.9	12.9	27.9
5	48.8	16.9	39.8	64.2	26.4	26.4	32.8	85.1	161.7	148.8	58.2	60.2	24.9	10.9	13.4	27.9
6	49.8	16.9	39.8	64.7	26.9	26.9	32.8	86.1	162.2	149.3	58.7	60.7	24.4	10.9	12.9	27.9
Average	**49.3**	**16.7**	**40.1**	**64.3**	**27.0**	**26.7**	**32.8**	**86.1**	**162.4**	**149.1**	**58.6**	**60.4**	**24.1**	**10.9**	**13.0**	**28.1**
SD	**0.5**	**0.5**	**0.6**	**0.4**	**0.4**	**0.3**	**0.2**	**0.7**	**0.7**	**0.5**	**0.4**	**0.3**	**0.8**	**0.2**	**0.4**	**0.4**
Error Coff.	**1.0**	**3.1**	**1.5**	**0.6**	**1.5**	**1.0**	**0.6**	**0.8**	**0.5**	**0.3**	**0.6**	**0.5**	**3.1**	**1.9**	**2.9**	**1.5**

Table 5: Measurements (mm) of the subject's face, from same day profile photographs.

Trial	t-ex	t-se	t-g	t-prn	t-pg	g-pg	prn-se	sn-prn	sto-sl	g-sn	sa-sba	ls-li	ls-sto	li-sto	sn-sto	∠ t-ex to horizontal
1	80.1	95.5	108.5	120.4	129.9	126.9	37.8	11.9	22.4	69.2	64.7	22.4	10.0	11.4	25.9	16.9
2	80.1	95.5	108.5	120.4	129.9	126.9	38.3	11.4	22.9	68.2	64.7	22.4	10.4	11.9	25.9	16.4
3	80.1	96.5	109.5	120.9	129.4	126.9	37.8	11.4	22.9	68.2	63.7	21.9	10.4	10.9	25.4	18.8
4	80.6	96.5	110.0	120.4	129.4	127.9	38.8	10.4	22.9	70.6	64.7	22.9	10.9	10.9	25.4	16.7
5	81.1	95.5	109.0	119.9	129.4	126.9	37.8	11.4	22.9	69.2	64.7	22.4	10.4	11.9	24.9	15.7
6	80.1	96.0	109.5	119.9	128.4	127.4	37.8	11.4	22.9	68.7	64.7	22.9	10.4	11.4	24.9	16.3
Average	**80.3**	**95.9**	**109.1**	**120.3**	**129.4**	**127.1**	**38.1**	**11.4**	**22.8**	**69.0**	**64.5**	**22.5**	**10.4**	**11.4**	**25.4**	**16.8**
SD	**0.4**	**0.5**	**0.6**	**0.4**	**0.5**	**0.4**	**0.4**	**0.5**	**0.2**	**0.9**	**0.4**	**0.4**	**0.3**	**0.4**	**0.4**	**1.1**
Error Coff.	**0.5**	**0.5**	**0.6**	**0.3**	**0.4**	**0.3**	**1.1**	**4.3**	**0.9**	**1.3**	**0.6**	**1.7**	**3.0**	**3.9**	**1.8**	**6.3**

Table 6: Measurements (mm) of the subject's face, from different day frontal photographs.

Trial	ch-ch	cph-cph	al-al	p-p	en-ex [r]	en-ex [l]	en-en	ex-ex	sa-sa	sba-sba	sa-sba [r]	sa-sba [l]	ls-li	ls-sto	li-sto	sn-sto
1	48.8	18.4	40.8	64.7	27.4	26.9	32.8	86.6	162.2	147.8	59.2	60.7	23.4	10.4	12.9	28.4
2	50.2	16.9	40.3	62.7	26.4	26.4	31.8	83.6	158.2	146.3	57.2	58.7	23.9	10.9	12.9	27.9
3	48.8	17.9	40.3	62.2	25.9	25.9	31.8	83.6	157.7	144.3	57.7	59.2	23.4	10.0	12.9	27.4
4	49.8	17.4	39.8	62.7	25.9	26.4	31.8	84.1	158.7	145.8	56.7	58.7	22.9	10.4	12.9	27.9
5	49.8	17.9	40.8	63.7	26.4	26.4	32.3	85.6	161.7	148.3	58.2	60.2	23.4	10.4	12.9	27.9
6	52.2	16.9	42.3	64.7	26.9	26.4	33.8	86.6	164.7	148.8	59.2	61.2	25.4	11.4	13.9	28.9
Average	**49.9**	**17.6**	**40.7**	**63.4**	**26.5**	**26.4**	**32.4**	**85.0**	**160.5**	**146.8**	**58.0**	**59.8**	**23.7**	**10.6**	**13.1**	**28.0**
SD	**1.3**	**0.6**	**0.9**	**1.1**	**0.6**	**0.3**	**0.8**	**1.4**	**2.8**	**1.7**	**1.0**	**1.1**	**0.9**	**0.5**	**0.4**	**0.5**
Error Coff.	**2.6**	**3.4**	**2.1**	**1.7**	**2.2**	**1.2**	**2.5**	**1.7**	**1.7**	**1.2**	**1.8**	**1.8**	**3.7**	**4.8**	**3.1**	**1.8**

Table 7: Measurements (mm) of the subject's face, from different day profile photographs.

Trial	t-ex	t-se	t-g	t-prn	t-pg	g-pg	prn-se	sn-pm	sto-sl	g-sn	sa-sba	ls-li	ls-sto	li-sto	sn-sto	∠ t-ex to horizontal
1	80.1	96.0	109.0	120.4	128.9	125.9	37.8	10.9	22.9	68.2	64.7	22.4	10.0	12.4	25.4	16.5
2	80.1	95.5	108.0	120.4	129.9	126.4	36.8	11.9	22.4	67.7	63.7	21.4	10.4	10.9	25.4	18.2
3	80.6	97.0	109.0	121.9	131.8	126.4	37.8	10.9	22.9	67.7	64.7	21.4	10.4	10.9	25.4	15.3
4	79.1	96.0	108.0	120.9	131.3	126.4	36.8	11.4	22.9	67.2	63.2	21.9	10.0	11.4	25.4	18.2
5	80.1	95.5	108.5	120.9	130.3	126.4	37.3	10.9	22.9	67.2	63.7	21.9	10.9	10.9	25.9	16.9
6	79.6	95.0	108.0	119.4	128.9	128.4	37.8	10.9	23.9	69.7	65.2	22.9	10.4	12.9	25.4	16.4
Average	**79.9**	**95.9**	**108.4**	**120.6**	**130.2**	**126.6**	**37.4**	**11.2**	**23.0**	**67.9**	**64.2**	**22.0**	**10.4**	**11.6**	**25.5**	**16.9**
SD	**0.5**	**0.7**	**0.5**	**0.8**	**1.2**	**0.9**	**0.5**	**0.4**	**0.5**	**0.9**	**0.8**	**0.6**	**0.4**	**0.9**	**0.2**	**1.1**
Error Coff.	**0.6**	**0.7**	**0.5**	**0.7**	**1.0**	**0.7**	**1.3**	**3.7**	**2.1**	**1.4**	**1.2**	**2.6**	**3.6**	**7.5**	**0.8**	**6.7**

Discussion

Overall the results indicate that highly repeatable photographs can be taken on the newly developed photography rig. However, this study also demonstrates that even with high standardization, some variations between photographs remain. It is, therefore, expected that between less standardized photographs variability is much greater, especially when subject-camera distances are not precisely controlled by systems like the projected light range finding mechanism described in this paper.

The range finding system described here successfully allows for non-contact, accurate, repeatable, and fast photography of subjects, while maintaining the natural head position. The non-contact properties of the rig are its primary advantages because: (i) subjects are recorded in their natural head position as they normally appear in life; (ii) it allows many subjects to be photographed in minimal time; and (iii) it makes photographing of young children easy and simple. These items cannot be achieved using contact systems, or other non-contact systems where various instruments may need to be placed about the head.

Also, the ability to take simultaneous profile and frontal photographs appears to be a significant advantage since facial measurements can be grossly adjusted for the affects of magnification by measuring distances of features from the focal plane on the opposite photograph. Although this method is not precise (since real distances cannot be calculated for measurements of features that do not fall in the same plane) it will compensate to a large degree for magnification distortions and increase the accuracy of photogrammetric measures. There are, however, some limitations with the photography rig as it presently exists and improvements can be made. For example, the flash units are not adjustable and hence lighting conditions may vary slightly between very tall and very short subjects. This is not favorable since differences may become significant during longitudinal studies of small children through to adulthood. This could be improved by having flash units adjustable according to the positioning of the horizontal camera beams.

Also, at present, the rig does not enable adjustment of the profile camera to obtain the same subject to lens distances for individuals of differing head widths. Therefore, a profile photograph of a person with a brachicephalic head will be magnified in comparison to person with similar features but with a dolicephalic head since subject to film plane distances differ. To enable more comparable profile photographs to be taken it would be necessary to mount a projected light range finding system on the profile camera also and slide the frontal camera accordingly so that the correct distances are achieved. However, such manual manipulations of the rig would significantly compromise subject throughput as they would need to be made for almost every individual.

Figure 11 also illustrates that care should be taken in subject positioning with the range finding system, since small deviations from the correct position results in much larger magnification errors on the photograph. For example, incorrect positioning of the subject with the light aligners apart by about 2mm results in a backward displacement of the subject by approximately 5 times as much (i.e., 10mm). An improvement may be made by placing the light pointers further apart so that they are more sensitive to subject movement. Additionally small variations in subject positioning may be caused by correctly positioning the light “v’s”, i.e., so they are touching, but with different degrees of overlap between photographs.

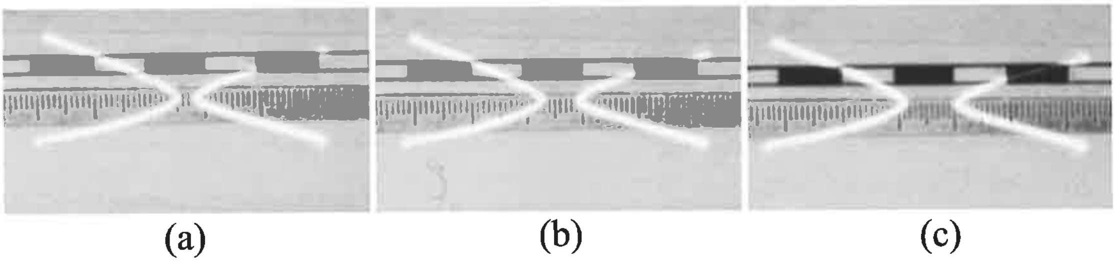


Figure 11: Examples of inter "v" distance at differing distances to the frontal camera **(a)** 10mm posterior to the point of sharpest focus; **(b)** 20mm posterior to the point of sharpest focus; and **(c)** 30mm posterior to the point of sharpest focus.

It should be noted that frequent calibration of the photography rig is not required, provided other uses of the rig do not tamper with the projected light range finders / camera position / zoom / focus. Tampering with equipment may be inevitable in a clinical setting where there are many users, however, it maybe minimized by restricting access to a few trained personnel.

Farkas *et al.* (1980) and Farkas (1994c) report photogrammetry to be valid (average difference between average photographic and direct measurements of no more than 1mm) for only a few measures including: n-sto; en-en; ps-pi; or-sci; n-sn; sn-c'; sn-sto; sto-li; cph-cph; and sbal-ls. Measures of the eye and mouth taken from photographs taken on this rig compare closely to direct anthropometric measures taken by Farkas *et al.* (1994a; 1994b) on similar (but not identical) populations, indicating that the scaling factor used here is adequate and that the rig is accurate (Table 8 and 9). Table 8 also shows comparisons to results obtained by other authors using photogrammetric methods to measure "Europeans". This table shows that the photography rig described here generally gives measures more representative of those obtained by Farkas et al. (1994a) on similar populations using direct anthropometric methods than other photography rigs. Moreover, measures from average faces (generated using photographs taken on the rig) also compare closely to Fakas et al. (1994a; 1994b) suggesting that the photographs taken on the rig described above are accurate (see Table 20).

Table 8: Comparison of average measurements made from photographs of Self-perceived Europeans (aged 18 to 34 years) taken on the photography rig to direct anthropometric measurements of "North American Caucasions" (aged 19 to 25) made by Farkas et al. (1994a; 1994b) and photogrammetric values obtained by other authors.

		Anthropometric			**Photogrammetric**											
		Farkas et al. (1994)			This study			Gavan et al. (1952)			Tanner and Weiner (1949)			Fraser and Pashayan (1970)		
	Sex	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
Mouth Width (ch-ch)	Male	54.5	3	109	55.2	3.1	27				49.9	4.2	65-70			
	Female	50.2	3.5	200	51.4	3.2	55									
Interpupillary Distance (p-p)	Male	66.9	3.9	40	64.7	3.9	27									
	Female	62.6	3.6	40	61.6	3.3	55									
Nose Width (al-al)	Male	34.9	2.1	109	37.9	1.5	17				38	2.6	65-70			
	Female	31.4	2	200	33.7	2.6	44									
	mix/unknown							40	0	2				31.5	0.5	20 M, 30 F

Table 9: Comparison of average measurements made from photographs of Self-perceived Central/South East Asians (aged 18 to 34) taken on the photography rig to direct anthropometric measurements of "Chinese Subjects" (aged 18 years) made by Farkas et al. (Farkas *et al.*, 1994a; Farkas *et al.*, 1994b)

		Anthropometric (Farkas et al. 1994a)			**Photogrammetric**		
		Mean	SD	n	Mean	SD	n
Mouth Width (ch-ch)	Male	48.3	6.8	30	52.9	5.3	26
	Female	47.3	3.3	30	50.7	3.4	25
Nose Width (al-al)	Male	39.2	2.9	30	42.2	2.8	12
	Female	37.2	2.1	15	38.2	1.9	15

Mouth Width

This chapter forms the basis for a full-length paper that has been published in *The American Journal of Phsycial Anthropology*, 2003 (Appendix 2), and a technical note that has been published in the Journal of Forensic Sciences, 2003 (Appendix 3).

Introduction

There are three (subjective) guidelines commonly used in facial approximation for predicting mouth width: (1) that mouth width is equal to interpupillary distance and can be determined by dropping two perpendicular lines from the center of the pupils down to the level of the mouth (Fig. 12a) (Krogman, 1962); (2) that the distance between the medial borders of the iris is equal to the mouth width (Fig. 12b) (Prag and Neave, 1997); and (3) that the corners of the mouth correspond to the junction between the upper canine and first premolar on both sides (Fig. 12c) (Krogman, 1962).

The existence of multiple guidelines for predicting the same trait indicates a lack of soft to hard tissue relationship knowledge, for it is logically and practically impossible for all of these guidelines to be (on average) correct. Method 3 appears to be the most popular guideline for determining mouth width and is reported in many major facial approximation texts including Krogman (1962), Prag and Neave (1997), Gatliff (1984), and Taylor (2001a). Prag and Neave (1997) have also suggested that method 1 is somewhat inaccurate, but provide no evidence for their statement.

This study aims to determine the accuracy of the three traditional and commonly used methods for estimating actual mouth width (cheilion-cheilion), as described above, by using photogrammetric techniques. This study also aims to investigate other potential anatomical relationships that may be useful in determining the width of the mouth. Although the guideline that the corners of the mouth align with the centers of the eye sockets and the widest points on the chin (Gatliff and Snow, 1979) has also been used to estimate mouth width this guideline will not be evaluated here since it does not appear to be commonly cited and the determination of "the widest points of the chin" is problematic from frontal photographs (too subjective).

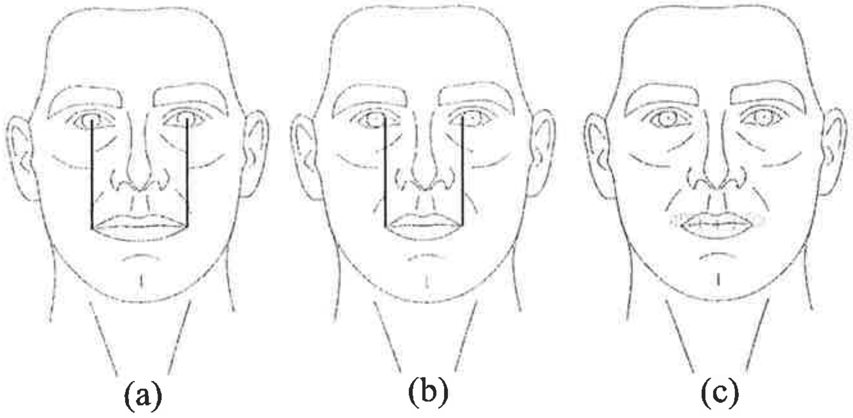


Figure 12: Traditional methods used to predict soft tissue mouth width
(a) cheilion determination from pupil centers
(b) cheilion determination from the medial aspects of the irises
(c) cheilion determination from the most lateral aspect of canines

Methods

Photography

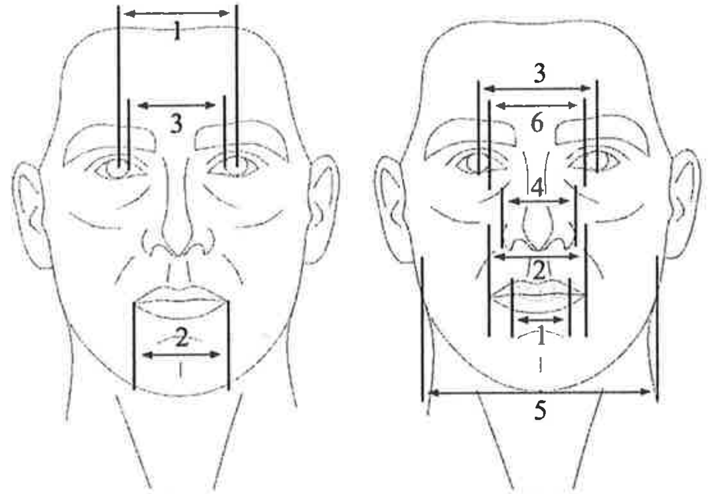
Photographs were taken of 72 Australians of European extraction (55 females, mean age 22.8 years, SD 4.3 years; and 27 males, mean age 21.4 years, SD 3.3 years), 51 Australians of East Asian extraction (25 females, mean age 20.6 years, SD 2.4 years; and 26 males, mean age 19.7 years, SD 2.2 years), and 13 Australian individuals of extraction from other populations (7 females, mean age 21.4 years, SD 5.7 years; and 6 males, mean age 19.5 years, SD 1.0 years), using the craniofacial photography rig described above. All participants were photographed in frontal and profile views (simultaneously), in both smiling and neutral expressions.

Photographs were taken using 200 ISO Ektachrome reversal slide film. Film development was standardized as much as possible by having the same qualified photographer develop the photographs in house. Once the slides were processed they were scanned into a computer using a Nikon® SF-2000 slide scanner. The resultant pictures were 1,200 pixels in width, 1,803 pixels in height and since they were large natural images they were saved in JPEG format for easier file management.

Mouth width guideline 1 & 2 (mouth width based on pupil width and medial iris border width)

Photographs were aligned so that the mid-plane, determined according to Farkas (1994a), was exactly vertical. Perpendiculars (vertical guides in Adobe® Photoshop® 6.0) were then dropped from the pupil centers, medial iris borders, and from the cheilions. Measurements were made for interpupillary distance, distance between the medial iris borders, and mouth width (Fig. 13a). Measurements (taken in mm) were converted to actual values using the scaling factor (4.975124) determined in the preceding chapter. Other magnification differences were corrected by adjusting measurements by 1% for every 10mm the feature being measured fell in front, or behind, glabella (the point of sharpest focus), as determined in the preceeding chapter. Since profile and frontal images were taken simultaneously, feature distances from the point of sharpest focus could be measured and, therefore, measurements could be roughly corrected for magnification.

Measurements were analyzed using histograms, F-tests, Pearson's correlations and two tailed t-tests (paired, two-sample equal variance, and two-sample unequal variance). Significance was set at $p<0.05$ but altered according to the Bonferroni adjustment (i.e., $p<0.02$).



(a) (b)

Figure 13: Main facial measurements taken
(a) Method 1 & 2:
1 – Inter-pupillary distance (p-p)
2 – Mouth width (ch-ch)
3 –distance between medial iris borders (iris-iris)
(b) Method 3:
1 – Distance between most lateral points of canines (c-c)
2 – Mouth width (ch-ch)
3 – Interpupillary distance (p-p)
4 – Nose width
5 – Jaw width at level of stomion
6 – Distance between most medial points of the iris (iris-iris)

Mouth width guideline 3 (mouth width based on canine width)

Photographs of participants used for this part of the experiment were the same as those who volunteered for method 1, however, since some individuals did not show their canine/premolar junction when the smiling photograph was taken sample sizes are less. Photographs of 61 European Australians (44 females, mean age 22.9 years, SD 4.4 years; and 17 males, mean age 21.6 years, SD 3.7 years), 27 East Asian Australians (15 females, mean age 21.1 years, SD 2.7 years; and 12 males, mean age 20.3 years, SD 2.8 years), and 5 Australian individuals of other populations (2 females, mean age 19.5 years, SD 0.7 years; and 3 males, mean age 20.0 years, SD 1.0 years) were used. Pairs of photographs (neutral and smiling) were superimposed, aligned based on the pupil centers, and rotated so that the mid-plane was exactly vertical.

Measurements were taken of interpupillary distance, mouth width, distance between the medial iris borders, jaw width at the level of stomion and nose width (Fig. 13b) and converted to actual values as described above. Inter-canine width (from most lateral points) was measured from the smiling photographs. Comparisons between smiling and non-smiling photographs were made by superimposition with alignment primarily based on the pupil centers. Measures were analyzed and compared using histograms, F-tests, Pearson's correlations, and two tailed t-tests (paired, two-sample equal variance, and two-sample unequal variance). Significance was initially set at the 95% confidence level but altered according to the Bonferroni adjustment (i.e., $p<0.01$).

Nose width and jaw width (Fig. 13b) were measured in addition to the other relevant measures to determine if they could be used to help predict mouth width. Least squares multivariate regression was used in an attempt to generate mouth width prediction equations. Independent variables were only included when their r^2 value was greater than 0.05. All statistics were calculated using the Microsoft Excel ® 2000 software package.

Results

Mouth width guideline 1 (mouth width based on pupil width)

Comparison of mouth width to interpupillary distance using two tailed paired t-tests showed a highly statistically significant difference ($p<0.001$) in all participant samples with mouth width being on average 17% smaller than interpupillary distance (Table 10). Therefore, the use of interpupillary distance to estimate mouth width is, on average, likely to cause an over estimation of about 11mm (sd 4mm) resulting in a mouth 6/5ths of its actual width.

In only 3% of the sample was the cheilion found beyond the border of the pupil on any side of the face. On average the cheilion fell approximately 5.9mm medial to the pupil in East Asian

Australians, and 4.7 mm medial in European Australians. Mouth width and interpupillary distance were correlated for both Central/South East Asians (females = 0.36, males = 0.22) and Europeans (females = 0.45, males = 0.37).

Average interpupillary distances were different at statistically significant levels between males and females for both populations of origin ($p<0.001$). However, mouth widths did not differ at statistically significant levels for Asians ($P>0.05$), while they did differ at statistically significant levels for Europeans ($P<0.001$). Average interpupillary distance and mouth width did not differ at statistically significant levels between the populations of origin ($P>0.05$).

Mouth width guideline 2 (mouth width based on iris width)

Comparisons between the distance between the medial iris borders and mouth width indicated an overall statistically significant difference ($p<0.001$). However, when the sample was separated by sex and population group, no statistically significant difference was found for female and male East Asians (Table 10), indicating that the guideline works fairly well for these groups but not so well for Europeans. Overall, the distance between the medial borders was found to underestimate mouth width by about 2mm (SD, 4mm), however, these two variables were correlated to some degree (Table 10).

Mouth width guideline 3 (mouth width based on canine width)

Comparisons of mouth width to canine width using two tailed paired t-tests showed a highly significant statistical difference ($p<0.001$) in all participant samples with canine width being on average 25% smaller than mouth width (Table 11). Therefore, the use of canine width to estimate mouth width is inaccurate and is likely to result in an underestimation of mouth width by about 13mm (SD, 3 mm), or a mouth 3/4ths of its actual width.

In every case (n=93) the cheilion was found lateral to canine/premolar junction. In East Asians this distance was about 5.9mm on average, and about 6.9mm on average in Europeans. Canine width and mouth width showed moderate correlations (Table 11), indicating, not surprisingly, that in both populations there is a trend for mouth width to increase, on average, with increasing canine width.

Table 10: Summary table of measurements made on participants in method 1 and 2 (all values in mm).

	Male Central/South East Asian (n=26)	**Female Central/South East Asian (n=25)**	**Male European (n=27)**	**Female European (n=55)**	**Other Individuals (n=13)**	**All groups (n=146)**
ch-ch	52.9	50.7	55.2	51.4	51.9	52.3
sd	5.3	3.4	3.1	3.2	3.5	4.0
p-p	65.5	62.2	64.7	61.6	61.3	62.9
sd	3.5	2.8	3.9	3.3	5.1	3.9
iris-iris	53.4	50.1	51.8	49.4	48.9	50.6
sd	4.3	2.6	3.7	3.1	4.7	3.9
Paired t-test for ch-ch to p-p	0.00	0.00	0.00	0.00	0.00	0.00
Paired t-test for ch-ch to iris-iris	0.70	0.41	0.00	0.00	0.01	0.00
Correlation for ch-ch to p-p	0.22	0.36	0.37	0.45	0.64	0.44
Correlation for ch-ch to iris-iris	0.18	0.30	0.34	0.45	0.70	0.40

Table 11: Summary table of measurements made on participants in method 3 (all values in mm).

	Male Central/South East Asian (n=12)	Female Central/South East Asian (n=15)	Male European (n=17)	Female European (n=44)	Other Individuals (n=5)	All groups (n=93)
ch-ch	54.2	51.2	55.0	51.4	52.5	52.5
sd	5.6	3.5	3.2	3.3	4.4	4.0
c-c	41.2	40.8	41.2	38.3	38.4	39.6
sd	1.5	1.9	1.5	1.7	3.2	2.2
(ch-ch - c-c)/(p-p - c-c)*100 (=a)	54.8	50.0	61.9	56.8	60.0	56.6
sd	24.2	15.7	12.5	12.6	9.6	15.0
(p-p - c-c)*a + c-c (=b)	54.5	51.3	55.2	51.6	52.7	52.6
sd	1.9	1.9	2.6	2.2	4.9	2.7
Average residual of b to ch-ch	0.2	0.1	0.1	0.1	-0.4	0.1
sd	5.8	3.1	2.9	2.9	2.2	3.4
Paired t-test for ch-ch to c-c	0.00	0.00	0.00	0.00	0.00	0.00
Paired t-test for ch-ch to b	0.90	0.91	0.89	0.80	0.87	0.78
Correlation for ch-ch to c-c	0.21	0.50	0.46	0.50	0.92	0.51
Correlation for ch-ch to b	0.08	0.46	0.53	0.49	0.89	0.52

Multivariate regression

Most regression equations generated for the variables represented in Figure 13b did not explain more than 50% of the variance of mouth width indicating that the variables measured were of limited use in predicting mouth width. As a result, this aspect of the study will not be further discussed.

A new guideline

No statistically significant difference was found between the measures of mouth width and canine width plus 57% of the cumulative distance between the lateral canine borders and the pupil centers on each side (value b, Table 11; and Fig. 14). This indicates that this measurement can be used to predict mouth width in the samples studied in this paper. The average difference between these measures was 0mm (SD 3mm), which is more accurate than results obtained using any of the three other traditional methods studied here.

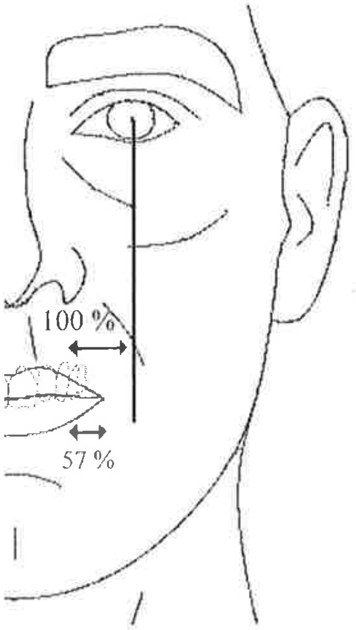


Figure 14: Example of mouth width prediction using the new suggested guideline
Assuming symmetry, cheilion falls at a point 57% of the way along the horizontal distance between the canine/first premolar junction and the pupil center. "One hundred percent" represents the horizontal distance from the canine/first premolar junction to the pupil center.

However, this new guideline, like guidelines 1 an 2, are limited because they rely not only directly on the skull but also on the soft tissue of the eye. In facial approximation the assumption is made that the eyeballs are centrally located within the orbits (Gatliff, 1984; Gatliff and Taylor, 2001; Krogman, 1962; Taylor, 2001a) allowing the use of average eye morphology to position other features such as the mouth. However, this assumption may be incorrect and may lead to further error. Some have attempted to evaluate the validity of the assumption (Eisenfeld *et al.*, 1975) but sample sizes have been small (n=9) and statistical power low. Consequently critical tests of the method are yet to be conducted. Results of Eisenfeld and colleagues (1975) suggest that actual interpupillary distance is over represented by central positioning in the orbits, however, it has been reported that that central positioning results in an under representation of actual interpupillary distance (Stephan, 2002). Further studies are needed to clarify the matter. However, it must be recognized that any inaccuracy in positioning/representation of the eyeballs is likely to result in inaccurate mouth width determination if the guidelines described above are followed. For this reason, instead of using "b" it would be better to predict mouth width as a percentage (>100%) of inter-canine distance.

Therefore, the ratio of inter-canine width to mouth width (chelion to chelion) was calculated using the 93 photographs of participants used to assess method 3. Inter-canine width was found to be equivalent to 75.8% of mouth width (or mouth width was about 133% of canine width). Even though canine width (c-c) and mouth width (ch-ch) differed at statistically significant levels between the European sexes and between European and Central/South East Asian females (see above), c-c to ch-ch ratios for all samples were fairly consistent.

When canine width was used as a percentage to estimate mouth width for the total sample (canine width/0.758) the average residual was –0.2 mm, s.d. 3.5 mm. Table 12 presents data across sub-samples when using the independent canine percentages and the 0.758 rule generally determined. It seems that the 75.8% inter-canine width rule worked least best for the "female Central/South East Asian" and the "other individual" groups suggesting that their independent ratios (79.9% and 73.1% respectively) may be of some value. However, none of the predicted mouth widths, determined using either the independent or the general percentage guidelines, differed from actual mouth widths at statistically significant levels (two tailed paired t-tests, $p>0.003$) indicating that the general guideline is sufficient.

These findings indicate that the general inter-canine width percentage guideline (0.758) predicts mouth width essentially as accurately as "b". Average error is barely more when using the inter-canine percentage guideline (-0.2mm, s.d. 3.5) than when using "b" (0.1mm, s.d. 3.4mm). However, the canine width percentage guideline is advantaged because unlike other guidelines it does not rely on subjective estimation of pupil location in the orbits. It, therefore, seems more logical to use the distance between the most lateral points of the canines as a percentage since guideline error is similar to that previously obtained and anatomical landmarks used for prediction are known. Since the 95% confidence range of the population mean for the c-c to ch-ch ratio (calculated from the sample mean reported in this study) is from 74.7% to 76.9% it seems valid to simply use 75% as the prediction rule, as opposed to 75.8%. This is useful since 75% is an even number that is easy to remember and apply in practical situations. The adjustment of the ratio by 0.8% slightly increased the inaccuracy of mouth width prediction in the sample reported here, but not by more than 0.6 mm on average for any of the groups studied.

Table 12: Summary table of canine width to mouth width ratios and mouth width prediction (all values in mm).

	Male Central/South East Asian (n=12)		Female Central/South East Asian (n=15)		Male European (n=17)		Female European (n=44)		Other Individuals (n=5)		All groups (n=93)	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
mouth width (ch-ch)	54.2	5.6	51.2	3.5	55.0	3.2	51.4	3.3	52.5	4.4	52.5	4.0
inter-canine width (c-c)	41.2	1.5	40.8	1.9	41.2	1.5	38.3	1.7	38.4	3.2	39.6	2.2
ratio c-c to ch-ch	76.6	8.2	79.9	5.0	75.2	4.1	74.7	4.4	73.1	2.6	75.8	5.3
mouth width estimation:												
c-c as a % of ch-ch (=c%)	53.7	1.9	51.0	2.4	54.9	2.0	51.3	2.3	52.4	4.4	52.3	2.9
Average residual of c% to ch-ch	-0.5	5.6	-0.2	3.1	-0.1	2.9	-0.1	2.9	0.0	1.7	-0.2	3.5
mouth width estimation: c-c/0.758	54.3	1.9	53.8	2.5	54.4	1.9	50.5	2.2	50.6	4.2		
Average residual of c-c/0.758 to ch-ch	0.1	5.6	2.6	3.1	-0.6	2.9	-0.9	2.9	-1.9	1.7		

Discussion

The result that mouth width differs largely, and at statistically significant levels, from both intercanine width and interpupillary distance strongly indicates that these measures are not accurate predictors of mouth width. Consequently, these guidelines should not be used in facial approximation to predict mouth width. It is therefore apparent that previous facial approximations constructed using these guidelines are likely to have incorrect mouth widths. This may have contributed to low recognition rates of facial approximations in experiments by Snow et al. (1970), Stephan and Henneberg (2001) and van Rensburg (1993).

Although the medial border of the irises approximates mouth width fairly well, on average, it remains different at statistically significant levels ($p<0.001$) being 2mm (SD, 4 mm) smaller than mouth width. However, this guideline is the most accurate of the three traditionally used facial approximation methods. Since the ratio guideline relies on bony points alone, unlike the medial iris guideline and the others described above, and is a fairly accurate predictor of mouth width (error of about 0.8mm on average) we suggest this new guideline should be used instead of the traditional methods. However, this guideline needs to be tested in other samples before it can be considered to be robust.

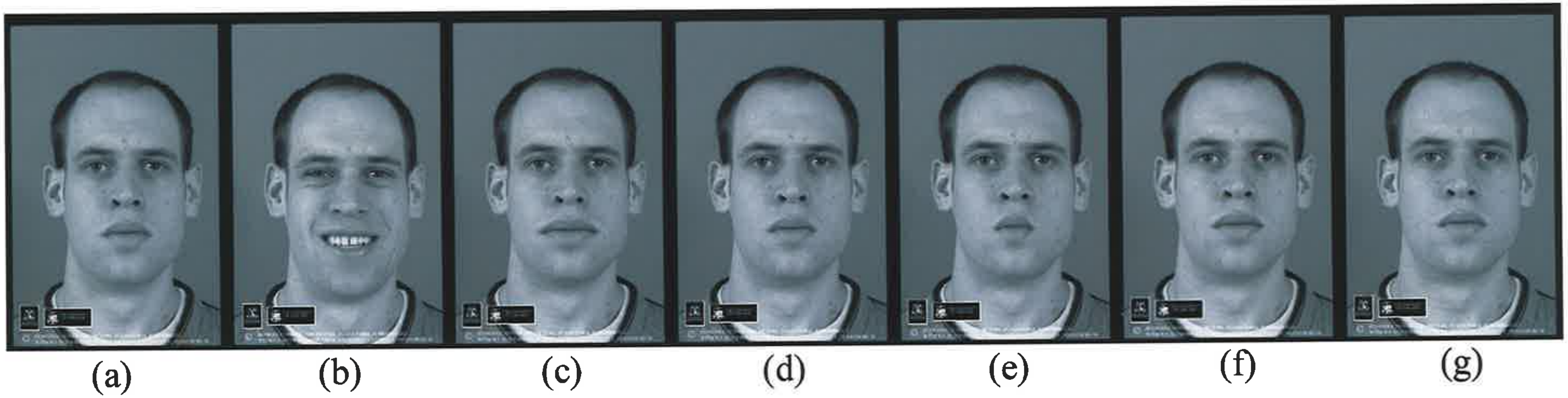


Figure 15: Examples of a face with mouth width determined according to various facial approximation guidelines
(a) Actual target face: normal relaxed; **(b)** Actual target face: normal smiling, to indicate position of upper canines; **(c)** Mouth width equal to distance between the pupils; **(d)** Mouth width equal to the medial iris borders; **(e)** Mouth width equal to distance between the canine/1st premolar junctions; **(f)** Mouth width equal to canine width plus 57% of the cumulative distance between the lateral canine borders and the pupil centers; and **(g)** Mouth width equal to canine width/75%.

Previously, the use of the traditional guidelines for determining mouth width of facial approximations is likely to have caused facial approximations to have incorrect mouth widths. This would appear to affect recognition as the errors introduced by using the pupils or the canines (the most popular method) are distinctly visible and appear to dramatically change the face (see Fig. 15). True positive recognition of these facial approximations is therefore expected to be more difficult, especially when combined with any soft tissue prediction inaccuracies of other features. However, the significance of mouth width by itself for facial recognition may not be substantial since Haig (1984) has found that mouth width inaccuracy alone has little effect on recognition rates. Haig's (1984) study altered mouth width in isolation so that lip and other facial morphology was exactly the same as the target individual. In facial approximation this is not the case as many features are likely to be inaccurately represented and may interact together to dramatically reduce the ability for the target individual to be recognized.

Although the new guidelines described here, are useful predictors of mouth width they do not give any indication of mouth position. Until more accurate ways of predicting asymmetry of the mouth have been determined, it is suggested that the mouth be placed symmetrically over the teeth unless skull morphology clearly suggests otherwise. Presently the vertical mouth location appears controversial as several authors provide differing opinions. For example, Fedosyutkin and Nainys (1993) suggest that stomion falls at the inferior border of the upper central incisors, while George (1987) suggests that stomion is located 1/3 of the way up the upper central incisors for females and a ¼ of the way for males. This appears to be an important characteristic to determine reliably since it independently appears to significantly contribute to face recognition (people seem to be particularly sensitive to superior movement of the mouth) (Haig, 1984).

It could not be determined in this study if the mouth width was equal to the distance between the mandibular second premolars, as has been suggested by Fedosyutkin and Nainys (1993), since the second premolar was in the shadow of the lips in the photographs used in this study. However, a relationship to the 2nd premolars may be probable since these teeth appear to be positioned closer to cheilion than the canines.

The results reported in this study appear to be directly applicable to two-dimensional drawing approximations where a photographic image of the skull is directly used for the basis of the facial approximation. The application of photogrammetric results, which are two dimensional, to three-dimensional scenarios may be problematic due to perspective distortions inherent to the photographic process (Farkas, 1994b; Farkas *et al.*, 1980). However, the two-dimensional error in the results of this paper may be less since photogrammetric measures of the eyes and the mouth are reported to be rather accurate; see Farkas *et al.* (1980), and p. 61-62 of this thesis. Until this study has been repeated in three-dimensions, any problems of perspective may be avoided by photographing the three-dimensional approximation (using similar methods as described in this study) to double check mouth width. This would seem not to be too troublesome as three-dimensional approximations (at least clay ones) are usually photographed anyway. Ultimately, three dimensional facial approximations should be constructed with the use of skull to face video-superimposition so that the actual hard to soft tissue relationship can be visualized and checked as the soft tissue is added over the hard-tissue. While superimposition is routine when using two-dimensional methods (e.g., drawing or computer-generated approximations) it should also be the case with clay approximations. Video superimposition is also convenient since mouth width can then be evaluated/checked in two dimensions during the facial approximation process without the need for photography.

Pronasale Position and Nose Projection

This chapter forms the basis of a full-length paper that has been published in *The American Journal of Phsycial Anthropology*, 2003 (Appendix 4).

Introduction

Nose projection is another feature for which many subjective and untested soft tissue guidelines exist in facial approximation (see Fig. 16 or Table 1). Although the nose has been regarded to be of little import for face recognition in frontal view (Fraser and Parker, 1986; Haig, 1984; Haig, 1986) its contribution to recognition may increase in other views like profile (Bruce, 1988). As a result, the nose appears to be an important character to predict accurately in facial approximation.

In general terms, the nose forms the exterior portion of the respiratory tract (Marieb, 1995). It is an anatomical structure that is usually located midsagittally, inferior to the eyes but superior to the mouth. The upper portion of the soft tissue nose has underlying bone support from the nasal bones and the frontal processes of the maxilla. The middle soft tissue portion is supported by the septal cartilage, which divides anteriorly into the lateral cartilages (Clements, 1969). The lower soft tissue portion is supported by the greater and lesser alar cartilages, which appear to vary greatly in size and shape (Clements, 1969; Marieb, 1995). Each greater alar cartilage consists of: a lateral crus that supports the alar; and a medial crus that forms part of the columella, which is located between the nostrils. The columella is also supported superiorly by the cartilaginous nasal septum and posteriorly by the anterior nasal spine. The cartilaginous septum joins with the hard-tissue septum, which is formed superiorly by the perpendicular ethmoid plate and inferiorly by the vomer bone. The large portion of cartilage that supports the external nose makes it a notoriously difficult feature to predict from the skull alone since cartilage relationships are not evident, at least given knowledge at this stage.

Not surprisingly, it has been found that the breadth of the nose is usually wider than the apertura piriformis (Schultz, 1918) and that soft tissue nasal height is correlated with hard-tissue nasal height, as is the prominence of the nasal bones (Macho, 1986). It is reported that 3/5th of the nasal aperture predicts actual nose width better than other methods (Hoffman *et al.*, 1991) and that the shape of the tip of the nose is independent of both the profile line of the bridge of the nose and the direction of the nasal septum (Macho, 1989). Macho (1989) provides evidence that the profile line of external nose does not follow its underlying hard-tissue structures as suggested by Schultz (1918) but suggests that nasal bone prominence is important for determining nasal shape, a finding consistent with the results of Posen (1967). Others have suggested that dorsal nasal humps appear to occur more frequently in individuals with Class II occlusions (Chaconas, 1969; Clements, 1969; Robinson *et al.*, 1986). It has also been suggested that "straight skeletal face profiles" are associated with straight nasal bridges; "convex skeletal face profiles" with convex nasal bridges; "concave skeletal face profiles" with concave nasal bridges; and that ~86% of individual nose morphologies can be correctly classified using these relationships (Robinson *et al.*, 1986). Schultz (1918) indicates that the subnasal point is on average ~6.5mm from the lower border of the nasal septum, while others suggest it is only 1-2mm below the nasal spine (Fedosyutkin and Nainys, 1993). Clements (1969) also suggests that the shape of the alar cartilages will be influenced by hypertrophy of the muscles located within the nostrils (e.g., compressor and dilatator nasalis muscles).

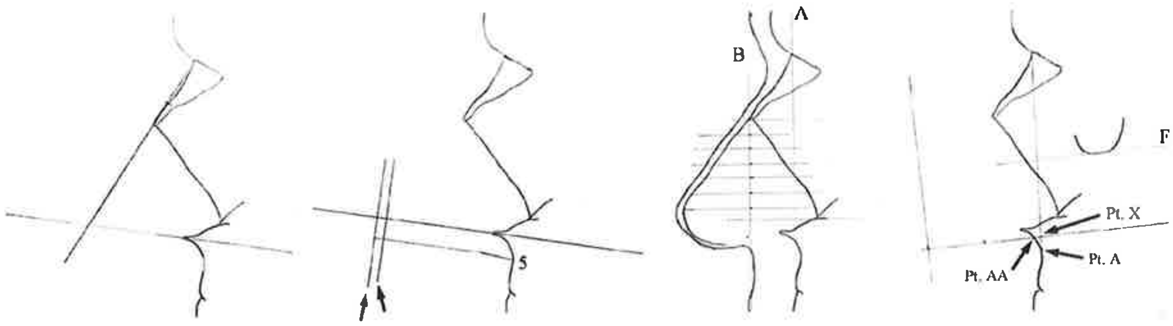
Several studies have examined soft- and hard-tissue nose projection and growth. Genecov and colleagues (1990) studied 64 "Caucasian" individuals longitudinally. They found that anterior nose projection continued to enlarge in males and females after skeletal growth had subsided.

In persons aged about 17 years they found nose projections from nasion to pronasale measured along the Frankfurt horizontal to be about 36mm (s.d. 3.5mm) for males and 34mm (s.d. 3.5mm) for females (Genecov *et al.*, 1990). The greater projection in males was found to be due to prolonged growth in comparison to females. Females had completed most of their growth by12 years with the nose only growing a further 2mm by age 17 years (Genecov *et al.*, 1990). In contrast, males showed a further 5mm of growth from 12 to 17 years of age (Genecov *et al.*, 1990). The angle of the nasal bone was also found to become more horizontal with age (Genecov *et al.*, 1990; Posen, 1967), changing by up to 6 degrees between ages 12 and 17 years (Genecov *et al.*, 1990).

Like Genecov *et al.* (1990), Nanda and colleagues (1990), in a longitudinal study of 40 "Caucasians", found females to complete nose growth earlier. They found nose projection to be ~90% complete in females at age 13 and ~70% in males (Nanda *et al.*, 1990). While female growth was almost complete at age 15, males continued to grow and appeared not to have stopped at 18 years (Nanda *et al.*, 1990). The mean projection values for females at 18 years of age, as measured from nasion to pronasale, perpendicular to the pterygomaxillary vertical plane, were ~29mm, while for males they were ~34mm (Nanda *et al.*, 1990). For nose height (nasion to pronasale, measured parallel to the pterygomaxillary vertical plane) females at age 18 averaged ~38mm while males averaged just over 39mm (Nanda *et al.*, 1990). Therefore studies by both Nanda *et al.* (1990) and Genecov *et al.* (1990) found sexual dimorphism (males larger than females) to be greater for projection of pronasale rather than pronasale height. Furthermore, both studies found only small increases (~4 degrees) in columella angle over the growth period, as measured either to the Frankfurt horizontal (age 7 to 17 years) (Genecov *et al.*, 1990) or the pterygomaxillary vertical plane (age 7 to 18 years) (Nanda *et al.*, 1990). Other authors (Behrents, 1985; Robinson *et al.*, 1986) have also made observations similar to those mentioned above.

With increasing adult age the nasal septum is reported to "sink" (Macho, 1989) with the nasal tip becoming more down-turned (Behrents, 1985; Posen, 1967; Subtelny, 1959) and the nose longer and wider (Behrents, 1985). The increasing vertical size of the nose may be due to the action of gravity on aged tissues that: are less hydrated; have decreased numbers of elastic fibers; and have atrophied collagen and muscle fibers (Patterson, 1980). Additionally, as supporting fibrous tissue is weakened with age the nasal cartilages separate, which may predispose the nose to the action of gravity (Janeke and Wright, 1971; Krmpotic-Nemanic *et al.*, 1971; Patterson, 1980). Some also report that the increasing size of the nose with age may be due to continued hyaline cartilage growth throughout life (Neave, 1998), however, this conclusion appears not to be supported by any published evidence. Most other hyaline cartilage appears to cease growth after the completion of puberty (Marieb, 1995).

Although several authors reported correlations between nose projection and the angle of the nasal bones (Chaconas, 1969; Robinson *et al.*, 1986) few systematic papers appear to have been published with specific regard to determining nose projection/pronasale position from the skull. There are, however, 4 published subjective methods that are commonly being used to predict nose projection (Fig. 16) and it is worth noting that all differ considerably.



Method 1 Method 2 Method 3 Method 4

Figure 16: Pronasale/nose-projection prediction according to four methods using a male adult skull of European extraction
For actual soft tissue to hard tissue relationship see Figure 19.

Method 1: Gerasimov (1971) method. Two tangents are projected, one following the last (distal) third of the nasal bones and the other following the general direction of the nasal spine. Where the tangents intersect indicates the tip of the nose (pronasale). When placing the nasal bone tangent priority was given to the superior surface of the nasal bones as would be done with real skulls where complete nasal bone profiles cannot usually be considered due to visualization limitations.

Method 2: Krogman (1962) method. Three times the length of the nasal spine (measured from junction of the vomer and maxilla to the tip of the nasal spine (Taylor, 2001a)) is added to the average soft tissue depth (according to Rhine and colleagues (Rhine and Campbell, 1980; Rhine and Moore, 1984) as suggested by Taylor (2001a)) at mid-philtrum (defined as a "midline point placed as high as possible before the curvature of the anterior nasal spine begins" (Taylor, 2001a)) following the general direction of the nasal spine (Taylor, 2001a). At the point of predicted nose projection a perpendicular was drawn to indicate the level of nose projection (black arrow) as indicated by Taylor (2001a). Also tested was a variation of this guideline by substituting the original nasal spine length for that determined by the margin of the most prominent lateral nasal aperture line to the tip of the nasal spine (nose estimation using this guideline indicated by gray arrow – see methods for further explanation).

Method 3: Prokopec/Ubelaker (2002) method. The nasal aperture is divided into 7 equal segments along a line at rhinion (line B) that is parallel to the nasion-prosthion plane (line A). The distance from line B to the edge of the aperture is measured and mirrored anteriorly. Two millimeters is added to each of these measures, the midline average soft tissue depths at nasion, mid-nasal bone, rhinion, and subnasale (according to Helmer (1984)) are added, and the points joined, allowing the profile shape of the nose and, hence, its projection to be estimated (Prokopec and Ubelaker, 2002).

Method 4: George (1987) method. The distance from nasion to point A (the point of most flexion on the maxilla in profile) is measured. A percentage of this distance (60.5% for males and 56% for females) is represented along a line parallel to the Frankfurt horizontal (line F) (George R., 2002, personal communication) beginning at point X, which falls on the nasion-point A plane at the height of point AA (a point located midway along the inferior slope of the anterior nasal spine). A line perpendicular to the Frankfurt horizontal is placed at the level of predicted nose projection.

It appears that the authors first proposing these techniques (or at least first publishing them) were: Gerasimov (1971) for method 1 and 3, Krogman (1962) for method 2, and George (1987) for method 4. Gerasimov (1971) originally proposed method 1 for determining pronasale position and nose projection while he used method 3 to determine the profile shape of the nose. However, following descriptions of Prokopec and Ubelaker (2002) it seems method 3 can also be used to indicate nose projection (see Fig. 16). Since Prokopec and Ubelaker (2002) provide a detailed description of method 3, in this study they will be linked with the technique, even though Gerasimov (1971) originally developed it. Methods 1 and 2 seem the most popular of the four methods, having been used by authors in many leading texts, e.g., method 1: Fedosyutkin and Nainys (1993), Gerasimov (1971), Prag and Neave (1997); and method 2: Gatliff (1984), Gatliff and Taylor (2001), Taylor (2001a). Despite the publication and broad use of all four methods of nose prediction, no accuracy/reliability tests have been published and it therefore appears that at least some practitioners have blindly followed informal observations of others. Although this is unavoidable in many instances, it is not favorable since these soft tissue prediction guidelines may be far from the truth.

Virchow (1912; 1924) apparently “emphasized that the prominence of the nose cannot a priori be predicted by the bony nose” (Macho, 1986; p.1392). However, Gerasimov (1971) p.54 reports that method 1 has been established as being reliable after “many years of work” and that this guideline will generally give the position of the tip of the nose, but no evidence is cited. Prag and Neave (1997), in contrast to Gerasimov (1971), have suggested that method 1 only approximates nose projection rather than determines it reliably, but again no evidence is provided. Stephan and Henneberg (2001) provide several examples in support for their statement that method 1 results in an over prediction of nose projection, and unrealistically large noses. It seems likely that method 4 may not be accurate in predicting actual nose projection (George R., 2002, personal communication) because the method is based on "aesthetic" surgical methods of Goode that describe a 'balanced' nasal projection when the nasofacial angle is at about 36° (Powell and Humphreys, 1984) which many individuals are not expected to possess naturally (note: "balanced" and "aesthetic" are subjectively loaded terms). However, in the three case examples presented by George (1987) method 4 appeared to work fairly well. Taylor (2001a) has indicated that method 3 does not have any sound basis in anatomy, yet uses method 2, which may also be questioned on similar grounds since no reliable and/or specific hard/soft tissue relationships, at the present time, are known to exist between the length of the nasal spine and the projection of the soft tissue of the nose.

While many authors (Fedosyutkin and Nainys, 1993; Gerasimov, 1971; Prag and Neave, 1997; Taylor, 2001a) suggest that the nasal spine indicates if the nose points up, down or is horizontal and use this relationship to predict nose projection/pronasale position, no formal tests have been published. While it may be correct in a general sense, using a direct relationship as in method 1 appears to be, at least, ambitious, given that most nose tips are located considerably above the nasal spine (as easily seen from lateral head radiographs).

The aim of this study is to test, using lateral head cephalograms, the accuracy of the four methods traditionally used in facial approximation to predict nose projection and/or pronasale position and improve them if possible.

Methods

The sample consists of 29 male (mean age 24 years, s.d. 10 years) and 30 female (mean age 23 years, s.d. 5 years) lateral head cephalograms of Australians of European extraction, randomly selected from the records of the Adelaide Dental Hospital and the University of Adelaide. While this sample may be biased somewhat toward patients thought to require orthodontic treatment, the sample also includes individuals who were not in need of treatment and who had been x-rayed for checkup. Since all individuals were drawn from the general population and may be subject to a forensic enquiry, no attempt was made to limit the sample to what may be considered "normal". That is, no individual was excluded if they displayed malocclusions, prognathism/retrognathism, missing teeth, and/or were pre/post-operative.

Five tracings were made for each lateral cephalogram. First, one soft tissue and one hard tissue tracing were made. Then, three duplications of the hard tissue tracing were made, giving four identical dependent hard tissue tracings and one soft tissue tracing in total for each lateral cephalogram. Tracings were made using a 0.5mm HB retractable pencil, and a fluorescent light box in a darkened room with aperture style shields to remove excess light generated by the light box. Each individual's tracings were marked with three identical reference points so that precise superimpositions could be made at a later time. The soft tissue tracing was isolated from the other four tracings, which were used to estimate nose projection, under blind conditions, according to methods described in Figure 16.

Tracing and measuring accuracies were assessed by comparing measurements of tracings and independent re-tracings of 10 individuals. We chose to examine measurements that would be used in the main experiment. The technical error of tracing/measurement was assessed as a coefficient of variation of the error (CVE). The CVE was calculated by taking the sum of the squared differences between test and retest and dividing it by 2x the number of re-measured individuals. The square root of the result was taken and divided by the mean of the test/retest result of the first individual. Coefficients of variation of error (CVE) were low for measurements of the tracings indicating that the combined tracing and the measuring techniques were rather repeatable (Table 13).

Table 13: The technical error of tracing/measurement assessed as a coefficient of variation of the error (CVE).

	CVE (%)
x	1
y	1
Nasion to Rhinion	4
Nasion to Point A	2
Nasion to Point AA	1
Anterior tip of nasal spine to profile line of nasal aperture	9

Areas of greatest variation appeared to be in determining/tracing the orbit, external auditory meatus, the lower anterior profile line of the nasal aperture, and the shape of the inferior aspect of the nasal spine. Figure 17 shows superimpositions of two trace/retraces that indicate the extremes of repeatability observed. Over the 10 independent retraces, Frankfurt horizontal repeatability was fairly high with the difference between tracings being on average 0 degrees with a standard deviation of 1 degree.

A variation of method 2 (plus three times the distance from the tip of the nasal spine to the border of the nasal aperture at its base) was included since on lateral cephalograms the junction of the vomer with the maxilla can be indistinct (Fig. 18) and it was thought that the aperture borders probably fall close to the vomer/maxilla junction. However, there is no existing evidence for this relationship. Although the outline of the vomer bone could be determined on some radiographs, it was generally difficult, and the precise point of the vomer/maxilla junction was hard to indicate since the transition between the two bones was smooth, showing little shape or density differences in profile. Figure 18 shows the delineation of the nasal septum, and the relative position of the vomer/maxilla junction on an x-ray of a dry skull. The anterior aspect of the perpendicular ethmoid plate was generally difficult to determine in the lateral head radiographs due to hard- and soft tissue shielding, as was the upper portion of the vomer bone. However, the distal portion of the vomer bone could be determined on 44% of cephalograms and when it was, the length of the nasal spine to the vomer/maxilla junction was estimated. This gave a sample size of 14 females (mean age 23 years, s.d. 5 years) and 12 males (mean age 28 years, s.d. 15 years, range 19 to 72 years). The border of the nasal aperture was also used for measuring nasal spine length in these and all other cephalograms.

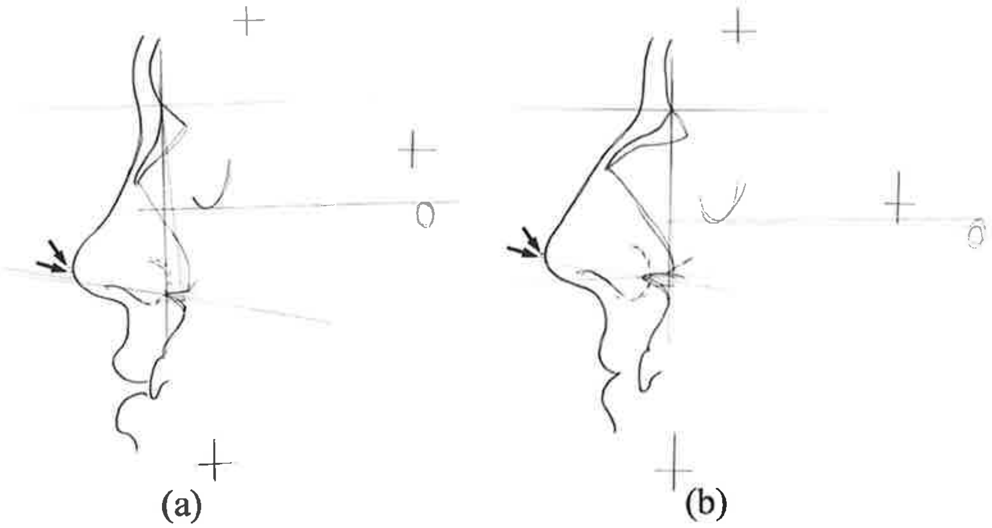


Figure 17: Selected superimpositions of independent tracings made from the same radiograph Note the placement of pronasale in each case indicated by the dash and highlighted by the arrows. **(a)** shows high repeatability superimpositions, although there is some difference in the inferior aspect of the nasal bone and inferior aspect of nasal spine curve. **(b)** shows low repeatability superimpositions, with differences between the representation of the external auditory meatus, orbit margin, inferior aspect of nasal spine curve, and mid portion of nasal aperture profile.

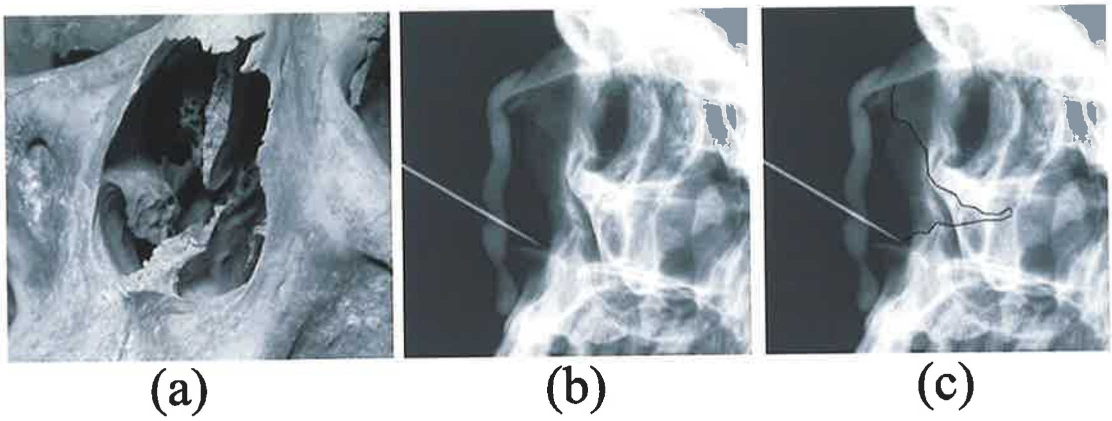


Figure 18: Nasal aperture and nasal septum delineation in lateral x-ray of dry skull **(a)** shows an oblique photographic image of the nasal aperture, vomer bone and perpendicular ethomoid plate. **(b)** shows the lateral x-ray of the skull with a pin indicating vomer/maxilla junction held in place by Bostik Blu Tak® **(c)** indicates the anterior profile line of the nasal septum (vomer bone and perpendicular ethmoid plate).

Once the soft tissue nose projection had been estimated on the hard-tissue tracings, the actual soft tissue tracings were superimposed. Measurements were made for actual and predicted nose projection/pronasale position using a Cartesian axis set up about the Frankfurt horizontal (x) and a perpendicular at nasion (y) (Fig. 19). The most anterior point of the nose (i.e., pronasale/nose projection) was determined by bisecting the curve of the anterior most tip of the nose as determined by sliding a perpendicular anteriorly along the plane of the Frankfurt horizontal. Additional features to those used in estimating nose projection/pronasale were also measured to determine if they explained any variance in pronasale position, these included: the length of the nasal bones from nasion to rhinion; and the angle of the nasal bones from nasion to rhinion with respect to the Frankfurt horizontal (Fig. 19). The angle of the nasal spine and that of the soft tissue nasal septum was also measured with respect to the Frankfurt horizontal (Fig. 19). All distance measures, but not angles, were reduced by a factor of 0.088 to obtain actual values since the cephalogram images were magnified by this factor in comparison to life.

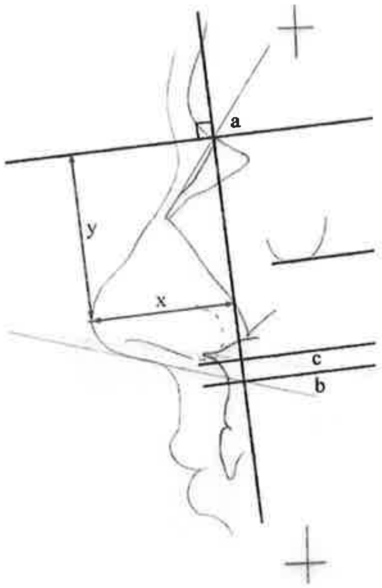


Figure 19: Measurement of pronasale/nose projection and additional angles
Black lines indicate those in the same plane as the Frankfurt horizontal or perpendicular to it. "x" represents pronasale projection; "y" represents pronasale "height"; "a" indicates nasal bone angle; "b" indicates the soft tissue nasal septum angle; and "c" indicates the nasal spine angle.

The predicted and actual measures of nose projection/pronasale position were compared statistically using paired and unpaired, two tailed, t-tests with significance levels set at $p<0.05$, but altered according to the Bonferroni adjustment where necessary. Other measures were analyzed using scatterplots and regression to try to determine more accurate ways of predicting nose projection and pronasale position. All data analyses were conducted using Microsoft Excel 2000®.

Results

Actual pronasale position averaged: in males, $x = 30.9 \pm 4.7$mm, $y = 44.3 \pm 3.5$mm; and in females, $x = 28.1 \pm 4.1$mm, $y = 43.0 \pm 3.5$mm. Nose projection (x) was significantly different between the sexes ($p<0.05$) while pronasale height (y) did not differ at statistically significant levels ($p>0.05$).

Table 14 summarizes the accuracy of methods 1, (the variation of) 2, 3, and 4. Table 15 summarizes the accuracies of method 2 using the vomer/maxilla junction instead of the profile line of the nasal aperture. Overall, Method 4 (George) and 3 (Prokopec/Ubelaker) performed best (error of about 2 ± 4mm) with method 4 having slightly lower standard deviations of error than method 3 (Table 14). The other methods performed much worse having either large average errors and/or large standard deviations of error. Method 1 (Gerasimov) performed worst of all having an average error of about 5mm and a standard deviation of error equal to ~9mm (Tables 14). Overall the methods predicted nose projection slightly better in males as estimated values were less often statistically different from actual values, and had lower mean errors than females (Table 14).

Table 14: Predicted nose projection compared to actual nose projection (mm).

Males (n=29)

	Actual	**Gerasimov (1)**	**Krogman Variation (2)**	**Prokopec (3)**	**George (4)**
X Average	30.9	37.2*	29.0	32.3	29.5
SD	4.7	11.6	10.8	6.5	4.2
X Error		6.2	-1.9	1.4	-1.5
SD		9.5	7.6	3.9	3.5

Females (n=30)

	Actual	**Gerasimov (1)**	**Krogman Variation (2)**	**Prokopec (3)**	**George (4)**
X Average	28.1	32.4*	24.0*	30.3*	25.3*
SD	4.1	10.8	8.8	5.1	4.5
X Error		4.3	-4.1	2.2	-2.8
SD		8.4	7.0	3.9	2.4

* indicates statistically significant difference from actual values, $p<0.01$ (equivalent to $p<0.05$ after Bonferroni adjustment for 5 tests)

Table 15: Predicted nose projection compared to actual nose projection (mm)

Male sub-sample (n=12)

	Actual	Krogman (2)
X Average	31.1	25.4
SD	5.0	9.2
X Error		-5.7
SD		6.4

Female sub-sample (n=14)

	Actual	Krogman (2)
X Average	28.7	24.4
SD	3.8	6.5
X Error		-4.3
SD		6.0

* indicates statistically significant difference from actual values, $p<0.05$

Table 16 summarizes accuracies of methods 1 and 3 used to estimate pronasale position. Accuracy for method 3 (Prokopec/Ubelaker) was modest having an error (shortest distance between the predicted position of pronasale and the actual position) of about 5mm with a standard deviation of about 2mm (Table 16). Method 1 (Gerasimov) performed poorly having an average error of about 11mm with a standard deviation of about 8mm. Overall, most predicted averages were found to differ at statistically significant levels from actual averages (Tables 14 and 16).

Table 16: Predicted pronasale position compared to actual pronasale position (mm)
Note that x values for methods (1) and (3) are repeated from Table 14.

Males (n=29)

	Actual	**Gerasimov (1)**	**Prokopec (3)**	**Regression**
X Average	30.9	37.2*	32.3	31.1
SD	4.7	11.6	6.5	3.7
X Error		6.2	1.4	0.2
SD		9.5	3.9	2.7
Y Average	44.2	51.7†	45.8†	44.7
SD	3.5	7.7	3.4	2.8
Y Error		7.5	1.6	0.4
SD		6.4	2.6	2.3
Average shortest distance		11.8	4.6	2.3
SD		9.0	1.9	1.7

Females (n=30)

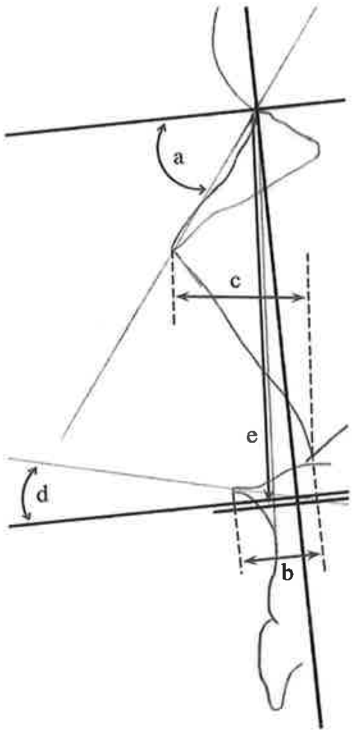
	Actual	**Gerasimov (1)**	**Prokopec (3)**	**Regression**
X Average	28.1	32.4*	30.3	28.0
SD	4.1	10.8	5.1	3.1
X Error		4.3	2.2	-0.1
SD		8.4	3.9	2.6
Y Average	43.0	49.7†	43.9†	42.3
SD	3.5	6.7	3.5	3.1
Y Error		6.7	1.0	-0.3
SD		4.9	2.1	2.0
Average shortest distance		10.5	4.4	2.6
SD		6.7	2.2	1.6

* indicates statistically significant difference from actual values, $p<0.01$ (equivalent to $p<0.05$ after Bonferroni adjustment for 5 tests)

† indicates statistically significant difference from actual values, $p<0.017$ (equivalent to $p<0.05$ after Bonferroni adjustment for 3 tests)

Nasal bone angle, measured between a line from nasion to rhinion and the Frankfurt horizontal, was not related to soft tissue nose height measured vertically from pronasale to the level of nasion (males $r^2 = 0.00$, females $r^2 = 0.08$), but was related to nose projection (males $r^2 = 0.52$, females $r^2 = 0.54$). These relationships were slightly stronger than for those between nose projection and the angle formed by the distal 1/3rd of the nasal bones (males $r^2 = 0.32$, females $r^2 = 0.51$). The distance between the tip of the nasal spine and the border of the nasal aperture at its base also explained some of the variance in nose projection in both males ($r^2 =$

0.35) and females ($r^2 = 0.15$). For males, the horizontal distance from rhinion to the most posterior point on the nasal aperture profile was related to nose projection ($r^2 = 0.20$). Using these three variables for males ($r^2 = 0.66$) and 2 for females ($r^2 = 0.58$) regression equations were generated (Fig. 20). For the sample studied, the regression equations were found to predict nose projection better than the four traditional methods listed above and estimated values did not differ from actual values at statistically significant levels (average error for both males and females = 0.0 ± 2.7mm, Table 16).



	r^2	S.E.
Pronasale projection:		
Males $x = -0.32(a) + 0.85(b) - 0.42(c) + 49.58$	**0.66**	**2.80**
Females $x = -0.41(a) + 0.37(b) + 49.87$	**0.58**	**2.70**
Pronasale height:		
Males and females $y = (-0.002(d) + 0.83)*(e)$	**0.64**	**2.17**

Figure 20: Regression equations generated to predict pronasale position (x and y) in relation to nasion and the Frankfurt horizontal
Image illustrates variables used in the equations. Black lines indicate those in the same plane as the Frankfurt horizontal or perpendicular to it. **(a)** = nasal bone angle, as measured from nasion to rhinion, from the Frankfurt horizontal. **(b)** = distance from tip of nasal spine to border of nasal aperture (at its base). **(c)** = distance from rhinion to most posterior point on nasal aperture border, measured perpendicular to the nasion/prosthion plane (see Line A, method 3, Fig. 16). **(d)** = nasal spine angle, as measured from the Frankfurt horizontal (positive if above Frankfurt horizontal; negative if below). **(e)** = distance of point X from nasion (also see Fig.16, method 4). All linear measurements measured in millimeters at life size (i.e., radiographic measures rescaled).

It was also found that the ratio of vertical soft tissue pronasale height (pronasale to nasion (y)) to hard tissue nasal height (Point X to nasion) did not differ substantially being on average 0.82 ± 0.05 in males and 0.84 ± 0.04 in females. Nasal spine angle (as measured from the Frankfurt horizontal) was found to explain some of the variance in the ratio (males $r^2 = 0.26$, females $r^2 = 0.22$) and when regression equations were generated (Fig. 20) they predicted nasal height in this sample rather accurately with estimated values not differing from actual values at statistically significant levels (average error for males = 0.4 ± 2.3mm, average error for females = -0.3 ± 2.0mm, see Table 16). Despite some relationship to soft tissue nasal height (see above), nasal spine angle was hardly related to the general direction of the columella (males $r^2 = 0.07$, females $r^2 = 0.05$).

When the two regression equations generated in this study were used to position pronasale from the skulls in this sample the average shortest distance between the predicted and actual position (males = 2.3 ± 1.7mm, females = 2.6 ± 1.6mm) was much less and more accurate than that given by the traditional methods (method 1: males = 11.8 ± 9.0mm, females = 10.5 ± 6.7mm; method 3: males = 4.6 ± 1.9mm, females = 4.4 ± 2.2mm, Table 16).

Discussion

Irrespective of the planes used for measuring, the average sex specific values of nose projection (x) tended to be ~4mm smaller in this study in comparison to others (Genecov *et al.*, 1990; Nanda *et al.*, 1990). However, nose height values (y) were ~5mm larger in comparison to those of Nanda and colleagues (1990). The same sexual dimorphism pattern was found in this study as others with males having larger values than females (Genecov *et al.*, 1990; Nanda *et al.*, 1990; Posen, 1967; Subtelny, 1959). Consistent with the findings of others (Genecov *et al.*, 1990; Nanda *et al.*, 1990), this study found differences between the sexes to be about two times greater for nose projection (x) than nose height (y).

Method 1: Gerasimov technique

This technique was highly unreliable, performing the worst of all four methods, and often resulted in overestimation of nose projection as has been suggested previously (Stephan and Henneberg, 2001). This is not surprising since the technique is rather subjective and imprecise because it relies upon the "general directions" of two bones, yet many facial approximation practitioners appear to regard this description as being precise enough for accurate replication (e.g., Fedosyutkin and Nainys, 1993; Gerasimov, 1971; Kustar, 1999; Prag and Neave, 1997). It is suggested that this technique not be used to construct noses in future facial approximations.

Method 2: Krogman technique

Both the actual technique and the variation reported in this paper performed rather badly, being 3rd best out of 4 methods. Based on the evidence found in this sample, it seems that these methods should not be used in future facial approximations. However, this recommendation may be harsh because: (i) the determination of the vomer/maxilla junction may be more accurate on real skulls than in radiographs; and (ii) that the distance to the profile line of the nasal aperture does not indicate/corroborate any inaccuracy of the original guideline since the two measures may not be the same.

However, it is worth noting that the variation of method 2 performed better than the original technique although sample sizes are small for the latter (Tables 14 and 15). The difference between the original technique and the variation of method 2, although statistically significant ($p<0.05$), was only 3mm, indicating that the difference between the two measures of nasal spine length was only about 1mm since the measures are multiplied by 3. The difference of 1mm does not appear to be significant because the error in determining the vomer/maxilla junction on real skulls may be of equal magnitude. For example, the dots Taylor (2001a) uses to mark the vomer/maxilla junction are ~1mm in diameter, and measurements from the dots are not consistent, that is, not always from the dot's center or its distal/proximal border (see Taylor (2001a), p. 393). Therefore, this evidence seems to indicate that both method 2 and its variation should not be used in future facial approximations.

This study demonstrates a significant limitation of the Krogman method because it relies on multiplying the nasal spine length by 3, which magnifies any error by three times. Furthermore, the use of the mid-philtrum soft tissue depth seems illogical when it is not directly related to the nose and more directly associated depths exist, like that at subnasale (see Helmer, 1984). In addition, placing the mid-philtrum depth so it follows the general direction of the nasal spine as indicated by Taylor (2001a) seems problematic since soft tissue mid-philtrum is actually located inferiorly to hard tissue mid-philtrum (George, 1993). From images provided by Taylor (2001a) and Gatliff and Taylor (2001), it can be seen that the mid-philtrum depth is often placed at the level of soft tissue subnasale, which is unrealistic. Using the mid-philtrum depth to represent subnasale is inaccurate since the subnasale depth appears to be about ~5mm greater than that reported by Rhine and Moore (1984) at mid-philtrum (Helmer, 1984). In Figure 12.25, p. 394, used by Taylor (2001a) to specifically describe the determination of nose projection the mid-philtrum marker actually falls well above subnasale (see Fig. 21), which clearly contradicts the directions of Rhine and Campbell (1980) for the placement of this soft tissue depth marker.

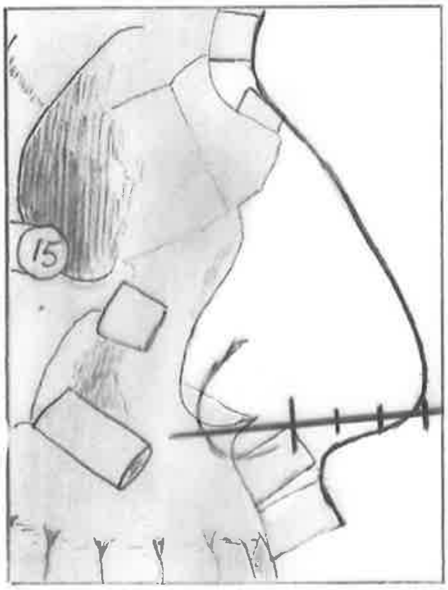
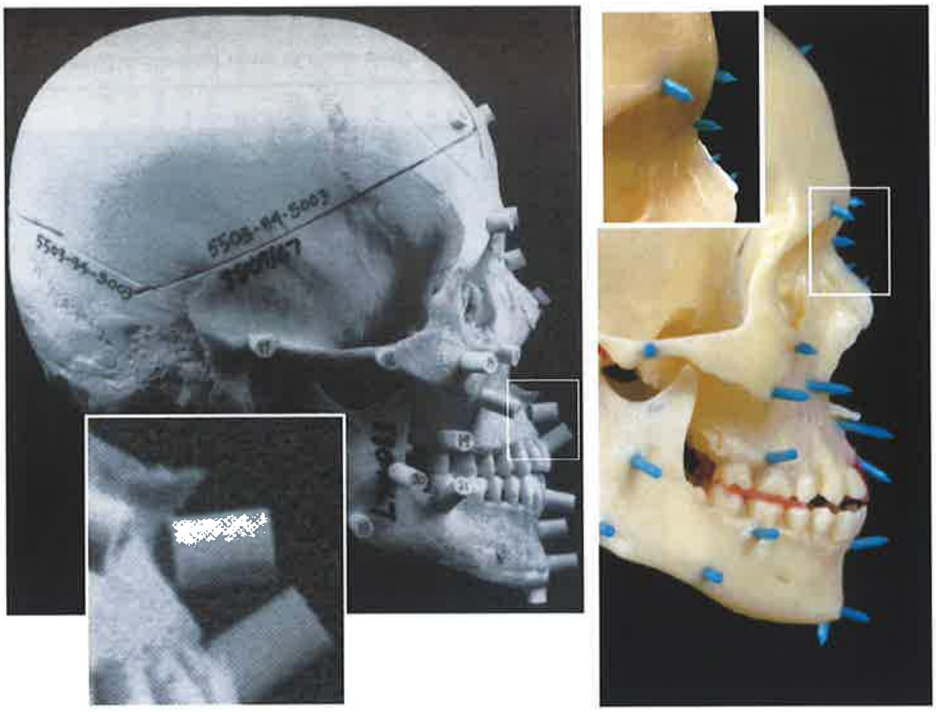


Figure 21: Determination of nose projection according to Taylor
Image drawn after Taylor (2001a) Fig. 12.25, p . 394. Note that the mid-philtrum soft tissue depth marker falls at subnasale and the upper lip margin marker falls at about mid-philtrum.

Other inaccuracies may also be introduced by the use of large cumbersome rubber cylinders to mark soft tissue depths as recommended by Taylor (2001a) and Gatliff *et al.* (Gatliff, 1984; Gatliff and Snow, 1979; Gatliff and Taylor, 2001), since they sometimes cannot be placed flush against the skull at particular anatomical locations, like nasion and subnasale (Fig. 22a), and because when glued to the skull they follow bone contours rather than actual directions of soft tissue measurement (see Fig. 22b). An alternative that seems more appropriate is to make a precise plaster or acrylic skull cast, and bore holes to properly locate small diameter, pointed, soft tissue markers (e.g., stainless steel nails or sharpened plastic/wood rods) placed at realistic angles (see Fig. 22b). Using a skull cast for the basis of clay facial approximation is also advantageous since the original can be referred to during the process (Taylor R personal communication) and fragile skulls are not put at risk of breaking due to the weight of the clay (Taylor, 2001a).



(a) (b)

Figure 22: Soft tissue depth markers
(a) use of large rubber cylinders. Inset highlights inaccurate representation. Image reproduced from Taylor (2001a) with reprint permission by K. Kinkade; **(b)** use of small diameter sharpened plastic rods inserted at realistic angles on a high precision skull cast.

Method 3: Prokopec/Ubelaker technique

This technique performed rather well, however, standard deviations of error were higher than that of method 4. While pronasale prediction was considerably more accurate than that of method 1, new methods described in this paper performed better. Therefore, this method should not be used if the regression equations described here, or method 4, can be. Despite the accuracy of this method in predicting pronasale position, observations made in this study suggest that much caution should be used when employing this method to predict profile nose shape as suggested by Prokopec and Ubelaker (2002) and Gerasimov (1971).

Method 4: George technique

This method estimated nose projection well being the best of the four methods studied. The success of this method seems to be attributable to the finding that nose projection did not vary considerably or consistently with nasion-point A height. Since this technique was based on considerably or consistently with nasion-point A height. Since this technique was based on methods to predict aesthetically pleasing nose projections it seems surprising that the method worked so well. The accuracy of this technique appears to indicate that either: natural nose projections are aesthetic for the majority of the population; or that the sample studied happened to possess "aesthetic noses" despite its random selection.

General discussion

Two out of the four traditional methods of facial approximation used to predict nose projection and/or pronasale position were inaccurate. It seems that new methods reported in this study predict nose dimensions better than the other two traditional methods that were quite accurate (methods 3 and 4). However, these regression equations need to be tested in other samples before they can be considered robust. It therefore seems that either the new methods reported here or the methods of George (1987) and Prokopec and Ubelaker (2002) can be used to estimate nose projection and/or pronasale position with a rather high degree of accuracy in future facial approximations.

When estimating nose projection from dry skulls caution may be needed if the distal ends of the nasal bones have changed shape as a result of dehydration (Taylor R., 2001, personal communication). We have noticed arching of the nasal bones in the coronal plane after processing and drying of skulls with long nasal bones (like those of koalas, possums, and kangaroos), but this phenomenon is yet to be systematically measured and verified in humans. Since human nasal bones are rather short and robust in comparison to other mammals the amount of dehydration related change may be minimal – however, the possibility that it may occur should not be disregarded. If it does cause significant changes then the soft tissue relationships of the hydrated nasal skeletal profile found in this study are probably different from those that exist between hydrated (living) soft tissues and dehydrated (dry) skulls.

The finding that the two most popular methods of determining nose projection/pronasale position are inaccurate suggests that many facial approximations have been made with inaccurate noses. It also suggests that practitioners using various methods will generally produce, on average, noses with different projections, e.g., method 1 leads to further projecting noses than method 2. Some support may be found for this in the literature. For example, profile images of facial approximations by Neave (Prag and Neave, 1997), who uses method 1, typically appear to have slightly more projecting noses than that typically represented on facial approximations by Gatliff (Gatliff and Taylor, 2001), who uses method 2. It is worth noting here that frontal images are of little use in evaluating nose projection because lighting shadows often produce misleading impressions (see for example Gatliff and Taylor (2001); Fig. 13.48, p.463). Given that popular methods of nose projection used in this study lead in some instances to extreme and even unrealistic estimations of nose projection in a sample of 59 individuals, it is expected that some facial approximations in the literature would also display unreal nose projections if guidelines were objectively followed. However, this appears not to be the case and suggests that either: facial approximations with unrealistic noses are not published; or facial approximation practitioners in some instances "curb" guidelines (maybe unconsciously) when the noses they predict seem unrealistic. If the latter is done, it seems reasonable given that facial approximation attempts to build a face as accurate as possible, however, one has to wonder why specific nose projection guidelines were ever used in the first place if they require subjective adjustments to be made ad libitum?

Eyeball Projection

This chapter formed the basis for a full-length paper publication in the *Journal of Forensic Sciences,* 2002 (Appendix 5).

Introduction

Besides the placement of average soft tissue pegs, one of the first procedures in any facial approximation is to locate the eyes within the orbit. Eyeball (or globe) positioning takes place in three planes: (i) the medio-lateral plane; (ii) the superio-inferior plane; and (iii) the anterio-posterior plane. Traditionally, globe location in the medial-lateral and superior-inferior planes has been accomplished by central positioning of the pupil (Gatliff, 1984; Gatliff and Taylor, 2001; Krogman, 1962; Taylor, 2001a), which can also be achieved according to methods of (Eisenfeld *et al.*, 1975). In the anterio-posterior plane, the globe has been placed by aligning the most anterior part of the cornea with an "imaginary" tangent from the superior to the inferior orbital rim (Gatliff, 1984; Gatliff and Taylor, 2001; Krogman, 1962; Taylor, 2001a) (Fig. 23).

Figure 23: Eyeball projection determined according to traditional methods

Despite the publication and promotion of the above positioning guidelines, there appear to be no published tests of these methods, by the original authors or by any other authors, in the scientific literature. It therefore appears that these guidelines have been based on untested observations and other facial approximation practitioners have followed the method blindly (e.g., Gatliff, 1984; Gatliff and Taylor, 2001; Stephan and Henneberg, 2001; Taylor, 2001a). Consequently, the accuracy and reliability of these facial approximation guidelines are unknown.

Experience of the author indicates that adherence to these globe positioning guidelines results in an under representation of globe projection and distance between the pupils. However, here only the projection of the globe in the anterio-posterior plane will be adressed. It seems worthy to note that while palpebral ligament attachments (Stewart, 1983) and the canthi of the eyelid (Angel and Krogman cited in (Caldwell, 1986)) may be useful to some extent for globe positioning in the medio-lateral plane, these features offer little use for determining globe position in the superio-inferior and anterio-posterior planes because these structures are not directly associated with the globe itself. If the palpebral ligament attachments and the locations of the canthi are used to position the eyeball in these two planes then unjustifiable assumptions must be made concerning curvatures of eyelid borders (for globe positioning in the superio-inferior plane) and eye proptosis (for anterio-posterior globe positioning).

Exophthalmometry, appears to be useful in determining globe position in the anteriror-posterior plane since it involves the measurement of the anterior protrusion of the globe in living subjects using standard instruments and methods. Exophthalmometry studies began as early as the 1870's (Emmert, 1870; Keyser, 1870), twenty-five years before scientific facial approximations began in 1895 (His, 1895b) so it is somewhat surprising that methods and knowledge have not been used previously with regard to facial approximation.

In exophthalmometry, globe projection measures are most commonly taken using a Hertel's, or Luedde's exophthalmometer (Fig. 24). Both exophthalmometers are used to measure the projection of the globe from the deepest point on the lateral orbital rim/s to the anterior most point of the cornea (Barretto and Mathog, 1999; Davanger, 1970; Drews, 1957).

When measuring, the exophthalmometer is placed firmly against the orbital rim and the projection of the cornea read off the scale. The lateral rim is chosen since it appears to have a thin covering layer of soft tissue regardless of the size or weight of the body (Knudtzon, 1949). However, using the lateral rim as a reference point is somewhat dubious because, in reality, it is not fixed (Davanger, 1970) (i.e., the lateral orbital rim position varies across individuals due to variation in skull growth; consequently a normal globe projection may, for example, be interpreted as being pathologically large if the lateral orbital rim is posteriorly displaced in comparison to the rest of the skull). Cohn (1865), who is reported to be the first to construct an exophthalmometer (Davanger, 1970), originally used the lateral orbital wall, however, he found it not to represent the "ideal plane" since it was often asymmetric between left and right sides and so he built a new exophthalmometer two years later that used the superior orbital margin as the reference (Cohn, 1867). However, others have reported that the supraorbital rim is just as variable as the lateral orbital rim (Keyser, 1900). Several other reference points have also been proposed (Drews, 1957), however, the lateral orbital wall is the most commonly used today because of the advantages listed above.

Although the accuracy of Hertel's Exophthalmometer has been challenged (Davanger, 1970), it is one of the most commonly used instruments to measure globe proptosis and has the advantages that: it is easy to operate; it measures both eyes simultaneously; and has cross hairs that allow correction for parallax (Quant and Woo, 1992). Despite the most likely source of error when using Hertel's exophthalmometer being misplacement of the instrument's foot plates (Davanger, 1970), both the Hertel's and Luedde's exophthalmometers are reported to have a measurement accuracy of about 1mm (Bertelsen, 1954; Drews, 1957; Musch *et al.*, 1985), with the lowest reported error being 0.5mm (Drews, 1957). However, the accuracy of the Hertel instrument starts to decrease if the transverse bar is not parallel to the frontal plane (usually caused by asymmetrical orbits) (Drews, 1957).

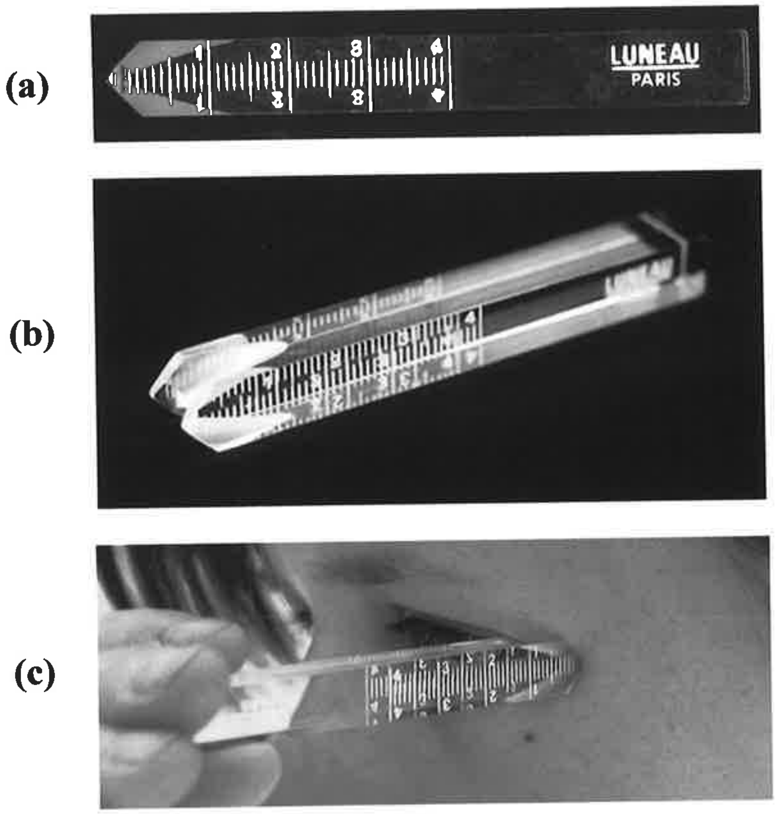


Figure 24: Example of an Luedde's exophthalmometer
(a) side view without parallax so double scale (one scale on each side of exophthalmometer) align along entire length of exophthalmometer; **(b)** superior oblique view showing thickness of exophthalmometer and double scale that is aligned for reading to reduce error as a result of parallax; **(c)** clinical use of exophthalmometer.

The aims of this study are: (i) to determine if there are any previously published papers that directly support, or refute, the facial approximation guideline for determining globe projection; and (ii) to determine if reported measures of globe projection from the lateral orbit correspond to the tangent from the superior and inferior orbital rims as predicted by the traditional facial approximation method.

Methods

A search of the ophthalmology and related literature was conducted for papers reporting globe projection values in normal healthy adults. Searches were conducted using Medline, Current Contents, and reference lists of exophthalmometry papers. Of those papers found, exophthalmometry values were collated in a database. Table 17 summarizes those papers and displays, where possible, sex specific averages, sample sizes, standard deviation and type of exophthalmometer used. Decimal places, as presented in reviewed articles, are presented in Table 17.

To establish if exophthalmometry values, as measured from the lateral orbital rim, correspond to the guideline used in facial approximation, measurements were also taken on crania from the deepest point of the lateral orbital rim to the tangent from the mid-superior to the mid-inferior orbital rim. Twenty-eight Caucasoid adult crania (as determined using standard osteological methods, e.g., sharpness of nasal sill; "pitching" of nasal bones; breadth of nasal aperture; shape of orbits; and degree of supraorbital ridge development etc. (see Briggs, 1998; Gill, 1998; Gill and Rhine, 1990; Krogman and Iscan, 1986; Stewart, 1979a; White, 2000) were used in the analysis. No attempt was made to separate this sample into sexes or ages because of its small size.

The mid orbital tangent was represented on the skulls using a plastic rod, fixed in the mid-sagittal plane of the left orbit using Bostik Blu-Tack®, placed over the plastic rod (Fig. 25a). A metal ruler was then used to measure the distance of the mid point of the rod from the deepest point on the lateral orbital rim (Fig. 25b/c). Before measuring, all skulls were inverted, but placed in a position equivalent to the natural head position (splanchnocranium rotated superiorly by ~5 degrees in comparison to the Frankfurt horizontal). The skull was also placed as symmetrically as possible according to methods of Drews (1957) using the intermaxillary suture, the foramen magnum, and the two glenoid fossae.

Table 17: Summary table of exophthalmometry studies found

Author (date published)	White American		Black American		Not Specified / Other		Exophthalmometer used
	Male	Female	Male	Female	Male	Female	
Bolanos-Gil-de Montes *et al.* (1999)					15.18 ± 2.16 (n=116)	14.82 ± 1.98 (n=185)	Hertel
Barretto and Mathog (1999)	17.00 ± 2.65 (n=34)	15.98 ± 2.22 (n=31)	18.23 ± 2.26 (n=33)	17.27 ± 1.44 (n=28)			Hertel and Luedde
Goldberg *et al.* (1999)					15.2 ± 2.8 (n=79)		MRI
Quant and Woo (1992)					16.66 ± 1.86 (n=120)	16.57 ± 1.78 (n=123)	Hertel
Dunsky (1992)			18.20 ± 2.97 (n=139)	17.46 ± 2.64 (n=170)			Hertel
Majekodunmi and Oluwole (1989)					13.5	15	Hertel
Bogren *et al.* (1986)	16 (n=53)		18 (n=47)				Hertel
Fledelius and Stubgaard (1986)					16.51 ± 2.26 (n=102)	16.01 ± 1.73 (n=101)	Rhodenstock apperatus
Migliori and Gladstone (1984)	16.51 ± 2.59 (n=127)	15.41 ± 2.34 (n=200)	18.49 ± 3.08 (n=113)	17.82 ± 2.57 (n=241)			Hertel
de Juan *et al.* (1980)	16.0 ± 2.30	14.7 ± 1.92	17.9 ± 2.86	17.1 ± 2.71			Luedde
Brown and Douglas (1975)	14.7 ± 1.7 (n=51)	17.0 ± 2.9 (n=87)					Hertel
Drescher-Benedict (1950) *					17.3 (n=100)		Hertel
Knudtzon (1949)	17.1 ± 2.08 (n=263)	16.8 ± 2.05 (n=99)					Hertel
Gormaz (1946)					14-16		Own
Soley (1942) †					15.9 (n=65)		Hertel
Ruedemann (1936)*					18.8 (n=1000)		Hertel
Wagener (1934)					18.0 (n=200)		Hertel
Lee (1930) †					14.4 (n=324)	14.8 (n=76)	Hertel
Jackson (1921)*					16-17 (n=4500)		Own
Helmbold (1916) †					16.67 (n=300)	15.68 (n=225)	Hertel
Woods (1915) †					12-14 (n=200)		Hertel
Birnbaum (1915) †					15 (n=120)	14.5 (n=30)	Hertel
Geraud (1912) †					13.6 (n=12)		Rollet-Durand
Birch-Hirschfeld (1900) †					14 (n=24)		Satler and Hering
Emmert (1870)					12-14 (n=200)		Own

The metal ruler was wide enough to be placed, simultaneously, on the lateral orbital rim and directly beside the rod (Fig. 25c). Measures were rounded to the nearest millimeter. This technique is similar to that employed when using Luedde's Exophthalmometer. Every effort was made to ensure the ruler was in the sagittal plane, as any deviation from it would introduce error, as would be the case if using Luedde's exophthalmometer (Davanger, 1970; Drews, 1957).

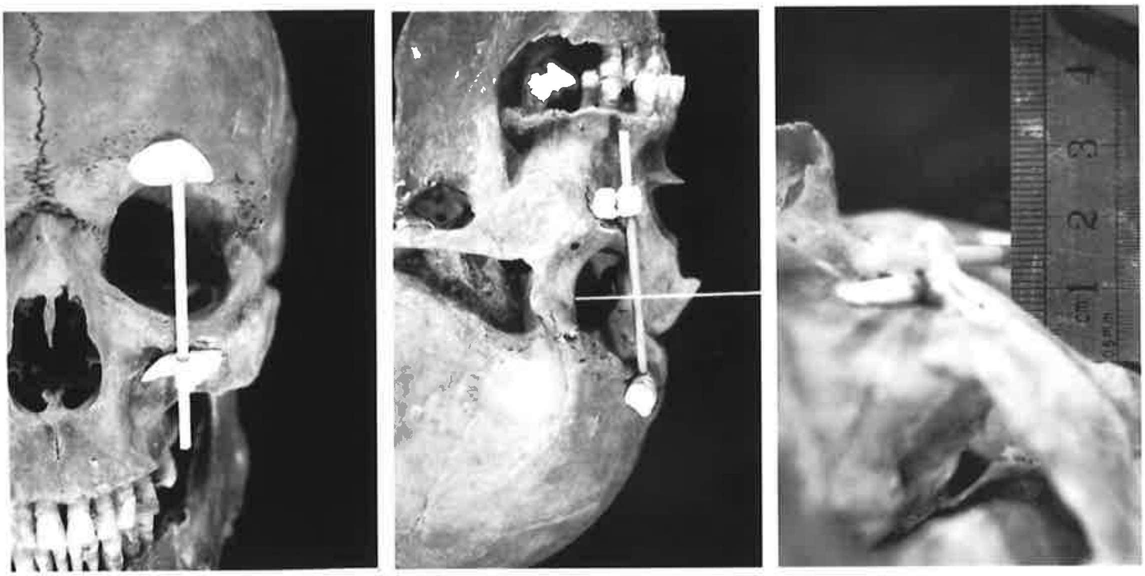

(a) **(b)** **(c)**

Figure 25: Eyeball projection determined according to the mid superior and mid inferior orbital margins
(a) Representation of the tangent (connecting the superior and inferior orbital rims in its sagittal plane) by plastic rod and held in place with Bostik Blu-Tack® placed over the rod. **(b)** measurement of the distance from the deepest portion of the lateral orbital rim to the tangent. **(c)** close up inferior-oblique view of the orbit showing the scale directly beside the rod and touching the lateral orbit while being held in the sagittal plane. In this case the projection of the tangent would be read as 13mm.

Exophthalmometry values were then compared to the tangent-orbit measures made. For exophthalmometry studies that were conducted on adult Caucasoid samples and reported, means, sample sizes, and standard deviations; two-sample t-tests were used to determine if statistically significant differences existed in comparison to the tangent-orbit measures made in this study. Significance was initially set at $p<0.05$ but altered according to the Bonferroni adjustment (i.e., since 11 tests were conducted, significance was taken at $p<0.0045$).

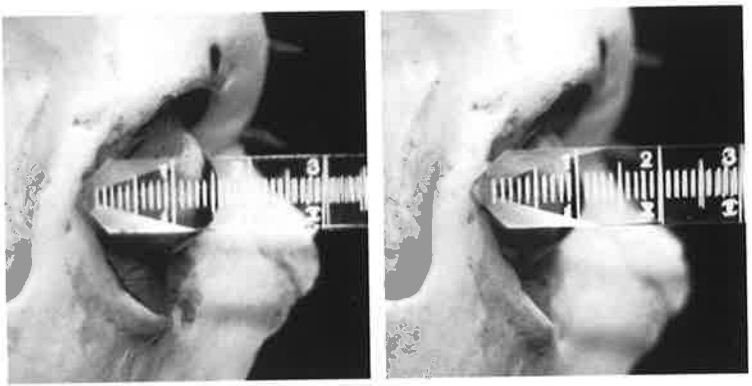
Results

A paper by Goldberg et al. (1999), using MRI techniques and a sample of 79 individuals, found that the anterior corneal surface falls, on average, 3.6 ± 3.3 mm anterior to the superior orbital rim and 11.3 ± 3.3 mm anterior to the inferior orbital rim. This observation shows that the facial approximation guideline of the tangent from the superior to inferior orbital rims under-represents actual globe projection in the vast majority of cases.

The average distance from the left lateral orbital rim to the tangent connecting the superior and inferior mid-sagittal orbital margins, as measured on 28 Caucasoid adult skulls, was 12.5 mm (SD 1.5 mm). This value was less than the average globe projection (16.2 mm, SD 2.3 mm) reported by the exophthalmometry literature (Table 18). Comparisons of tangent-orbit measures made in this study to reports of exophthalmometry in similar samples showed highly statistically significant differences ($p<0.006$) in every case (Table 18). This further supports the conclusion that the facial approximation guideline for anterior globe projection is inaccurate. The magnitude of this difference (4mm) also appears to be considerable in relation to other orbital measures like eye fissure length and height. Figure 26 illustrates the correct positioning of a prosthetic eyeball on a skull in comparison to the incorrect position obtained using traditional methods.

Table 18: Comparison of exophthalmometry measures taken on Caucasoid adults to the tangent-orbit measures made in this study using two-sample t-tests.

Study	sex	n	average	sd	t	p< normal (rescaled p according to Bonferroni's method)
Goldberg *et al.* (1999)	all	79	15.20	2.80	6.48	0.0005 (0.0055)
Barretto and Mathog (1999)	male	34	17.00	2.65	8.48	0.0005 (0.0055)
Barretto and Mathog (1999)	female	31	15.98	2.22	7.21	0.0005 (0.0055)
Fledelius and Stubgaard (1986)	male	102	16.51	2.26	11.16	0.0005 (0.0055)
Fledelius and Stubgaard (1986)	female	101	16.01	1.73	10.64	0.0005 (0.0055)
Migliori and Gladstone (1984)	male	127	16.51	2.59	11.05	0.0005 (0.0055)
Migliori and Gladstone (1984)	female	200	15.41	2.34	8.95	0.0005 (0.0055)
Brown and Douglas (1975)	male	51	14.70	1.70	6.06	0.0005 (0.0055)
Brown and Douglas (1975)	female	87	17.00	2.90	10.76	0.0005 (0.0055)
Knudtzon (1949)	male	263	17.10	2.08	14.76	0.0005 (0.0055)
Knudtzon (1949)	female	99	16.80	2.05	12.31	0.0005 (0.0055)
Total	all	1174	16.20	2.30		
tangent-orbit measures made in this study	all	28	12.43	1.53		



(a) **(b)**

Figure 26: Difference in globe position as a result of using either exophthalometry measures **(a)** or traditional facial approximation methods **(b)** on the same skull
A Luedde's exophthalmometer indicates anterior cornea distance from the lateral orbital rim in both cases. Note that camera position has changed with globe protrusion to ensure alignment of double scale of exopthalmometer.

Discussion

Previously published MRI data (Goldberg *et al.*, 1999), and measurements taken from skulls in this study in conjunction with published exophthalmometry studies (Table 17 & 18), indicate that the traditional facial approximation guideline for determining globe projection is inaccurate and is likely to underestimate the position of the cornea by more than 2.5mm (average underestimation = 3.7mm). It is, therefore, suggested that the traditional facial approximation guideline should be replaced by exophthalmometry values as measured from the lateral wall of the orbit.

Exophthalmometry values, as measured from the lateral orbital wall, appear to be useful to facial approximation practitioners since they have been comprehensively studied. Values have been calculated for numerous adult populations including: Chinese (Quant and Woo, 1992); Mexican (Bolanos Gil de Montes *et al.*, 1999); White American (Barretto and Mathog, 1999); Black American (Barretto and Mathog, 1999); and African (Majekodunmi and Oluwole, 1989). Studies also cover a relatively wide range of age groups. Fledelius and Stubgaard (1986) give values for children from 5 to 20 years, Nucci et al. (1989) give values for children aged 3 to 10 years, and Gerber et al. (1972) provide values from 10 to14 years of age. Although not shown in Table 17, many of the samples used by these studies have also been divided into age groups. Furthermore, globe projection has been shown to increase up until the late teenage years (about 17 years of age) when they reach adult values (Fledelius and Stubgaard, 1986). Values also appear to be slightly greater (1 to 2mm) for American Blacks than Whites (Table 17). Average male values appear to be consistently larger than females, however, the differences are generally not more than 1 mm (Jackson, 1921), being about the same as published instrument errors (Bertelsen, 1954; Drews, 1957; Musch *et al.*, 1985), indicating that this difference may not be real. It is also commonly reported that globe projection is larger on the right side (Drews, 1957; Gerber *et al.*, 1972; Knudtzon, 1949; Quant and Woo, 1992; Schlabs, 1915) but this asymmetry is not usually more than about 2mm in Whites (de Juan *et al.*, 1980; Dunsky, 1992; Fledelius and Stubgaard, 1986; Knudtzon, 1949; Migliori and Gladstone, 1984; Nucci *et al.*, 1989; Quant and Woo, 1992) and Mexicans (Bolanos Gil de Montes *et al.*, 1999), and about 3mm in Blacks (Bogren *et al.*, 1976; de Juan *et al.*, 1980). Asymmetries of 3-4 mm have, however, been recorded in normal healthy subjects (Bogren *et al.*, 1986; Majekodunmi and Oluwole, 1989). Majekodunmi and

Oluwole (1989), and Fledelius and Stubgaard (1986) report more proptosis of the left rather than the right side. However, other authors report that the difference between the sides is not statistically significant (Dunsky, 1992; Migliori and Gladstone, 1984).

In general, weak to no correlation in exophthalmometry values have been found for height (de Juan *et al.*, 1980; Majekodunmi and Oluwole, 1989; Migliori and Gladstone, 1984; Quant and Woo, 1992), head length (Quant and Woo, 1992), head width (Quant and Woo, 1992), temple width (Quant and Woo, 1992) and weight (de Juan *et al.*, 1980; Majekodunmi and Oluwole, 1989), but stronger correlations have been found for inter-orbital distance (Bertelsen, 1954; Quant and Woo, 1992) and corneal pituitary distance (Bogren *et al.*, 1986).

Bertelsen (1954) has suggested, from measurements he made, that proptosis can be predicted by adding 1mm to 15mm for every 4mm that pupillary distance increases beyond 61mm (and to remove 1mm from 15mm for every 4mm that the pupillary distance is less than 61mm), however, error rates are not reported. It has been suggested that globe projection increases with shallower orbits (Dunsky, 1992; Emmert, 1870; Migliori and Gladstone, 1984) and this has been proposed as a possible determinant of the higher exophthalmometry measures in blacks (Dunsky, 1992; Emmert, 1870; Migliori and Gladstone, 1984). The possibility exists that correlated measures may be useful in predicting individual values of globe projection, however, more research in this area is needed.

It seems illogical that some previous exophthalmometry studies report distances to two decimal places (see Table 17) considering that the error of measurement is about 0.5 to 1mm (Bertelsen, 1954; Drews, 1957; Musch *et al.*, 1985). It is, therefore, suggested that in facial approximation exophthalmometry values only be used to 0.5mm accuracy. It is also unlikely that eyeballs can be located beyond this precision in three-dimensional clay approximations anyway.

It may be considered by some that the inaccuracy of the traditional guideline, probably not much more than about 4mm on average, is quite small and may not be of significance if it does not affect facial approximation recognition. However, this appears not to be the case. Firstly, a difference of 4 mm is fairly large when dealing with small features of the face such as the eyes. For example, it is equal to ~13% of the eye length, en-ex, (Farkas *et al.*, 1994a) and ~37% of eye height, ps-pi, (Farkas *et al.*, 1994a). Secondly, errors are introduced into the facial approximation each time a feature is built. Therefore, this error will accumulate as many features are approximated and cause the final facial approximation to largely differ from the actual target individual, probably resulting in misidentifications (Stephan and Henneberg, 2001). Consequently, it seems important to limit the error introduced in all facial approximation guidelines to make facial approximations as exact and recognizable as possible.

The complete absence of papers referencing exophthalmometry studies in the facial approximation literature, and the use of a guideline that deviates from, and is unsupported by mainstream ophthalmology, are rather surprising. It indicates that facial approximation practitioners have blindly followed methods suggested by others, ignoring relevant exophthalmometry literature that uses accurate and reliable methods (Bertelsen, 1954; Drews, 1957; Musch *et al.*, 1985). This is surprising for two reasons: (i) a major priority of facial approximation practitioners would appear to be comprehensive and up to date knowledge and understanding of the anatomy and soft/hard tissue relationships of *all* facial features; and (ii) if facial approximation is really a blend of science and art as it is reported to be (Taylor, 2001a),"facial approximationists" should be conducting frequent literature reviews in an attempt to keep pace with new scientific knowledge and review it in a logical and scientific manner. It appears that the lack of a thorough literature review and/or logical assessment of it has, in this case, led to the use of an inaccurate technique for the last 40 years despite more reliable and accurate scientific methods that have existed for the last 80+ years. Data directly refuting the facial approximation guideline for globe projection have also been available since 1999 (Goldberg *et al.*, 1999).

Since this research on globe projection has been conducted and published, independent investigators have repeated the results obtained here using their own methods. Wilkinson and Mautner (2003) report from MRI measurements that the traditional globe projection guideline underestimates actual globe projection by 3.9mm, as compared to the average underestimation of 3.7mm reported in this thesis. E. Craig (2003) has suggested that the traditional globe projection guideline was not inaccurate because if it is used to position the iris, correct globe projection is achieved. However, it is clear in the literature that the traditional guideline described predicts the anterior cornea, not the iris; see Krogman (1962) p. 266, Krogman and Iscan (1986) p.429, Taylor (2001b) p. 381, Gatliff and Taylor (2001) p. 429, Gatliff and Snow (1979) p.29, or Gatliff (1984) p. 328.

Superciliare Position

This chapter formed the basis for a full-length paper publication in *Forensic Science International*, 2002 (Appendix 6).

Introduction

Another guideline that has been suggested for facial approximation is that superciliare (the most superior part of the eyebrow) is located directly above the lateral point of the iris (Taylor, 2001a) (Fig. 27). This guideline has also been used in the past to cosmetically position the arch of the female eyebrow when plucking or waxing (Campsie, 1997; Campsie,

1998) and a similar guideline has been used in plastic surgery with superciliare being aligned based on the lateral limbus, which lies near the lateral iris (Brennan, 1980; Ellenbogen, 1983; Powell and Humphreys, 1984). The guideline of superciliare being located above the most lateral point of the iris may be close to the truth since the arch of the eyebrow generally appears directly above the eye.

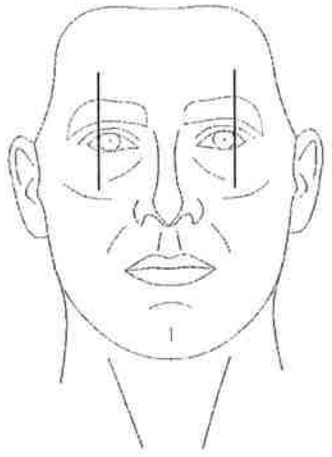


Figure 27: Guideline for determining superciliare from the lateral iris

Few studies have been conducted on the eyebrows. Rozprym (1934) has presented a general study of eyebrow morphology (and eyelashes) in over 500 individuals describing eyebrow form and providing classifications for them which appear to be useful. Rosprym (1934) found large variations in eyebrow form between individuals. Oestreicher and Hurwitz (1990) have also conducted studies on the position of the eyebrow, but did not specifically address superciliare position. They studied the right eyebrow on 46 males and 30 females (Oestreicher and Hurwitz, 1990). Their results indicated that mid eyebrow height tended to decrease with age with respect to the superior orbital rim, although the trend was not statistically significant ($p>0.05$). The difference between the young (<40 years) and old groups (>59 years) was only 0.71mm. However, females were found to have eyebrows, which were higher than males at statistically significant levels (males = -3.23mm, females = -0.99mm, "-" indicates brow position below the superior orbital rim, $p<0.004$).

This study aims to determine the accuracy and reliability of the traditional facial approximation guideline to position superciliare, both metrically and non-metrically, using photogrammetric methods.

Methods

One hundred and twenty eight participants, aged 18 to 30 years, average 21.4 years (SD 3.8 years), were photographed on the craniofacial rig described above. Frontal photographs of participants (in a relaxed, natural head position, with lips closed) were taken. Photographs were scanned into a computer using a Nikon® SF-2000 slide scanner and measured in Adobe® Photoshop® 6.0. The resultant pictures were 1,200 pixels in width, and 1,803 pixels in height.

All images were rotated as required so that the mid-saggital plane, as defined by Farkas (1994c), was exactly vertical. Superciliare was defined as the most readily determinable superior point of the eyebrow when all of the face could be seen using a Diamond View® 1995SL 483mm monitor. The horizontal distance from the lateral iris to superciliare was measured on each side of the face (Fig. 28). Four other measures were taken to determine if there was any relationship between them and the position of superciliare. Those measures were: the distance from the midline to the pupil center; the distance from midline to alare; the distance from midline to cheilion; and the vertical distance between the endocanthion and stomion (Fig. 28). All measures were adjusted by a scaling factor of 4.975124 (p. 52) to obtain actual values in millimeters.

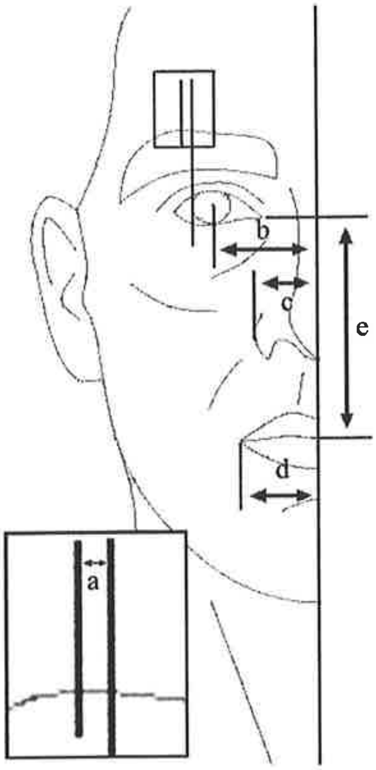


Figure 28: Measurements taken in this superciliare study
(a) distance of superciliare from lateral iris border, "a" was positive if it fell lateral to the lateral iris and negative if it fell medial; (b) the distance from the midline to the pupil center; (c) the distance from midline to alare; (d) the distance from midline to cheilion; (e) the vertical distance between the endocanthion and stomion.

Participants were also asked to indicate if they plucked/waxed their eyebrows and in which location they did so according to Figure 29. Although plucking in region 2 has the potential to dramatically alter the position of superciliare the decision was made not to exclude these individuals since they may also be subject of a forensic inquiry.

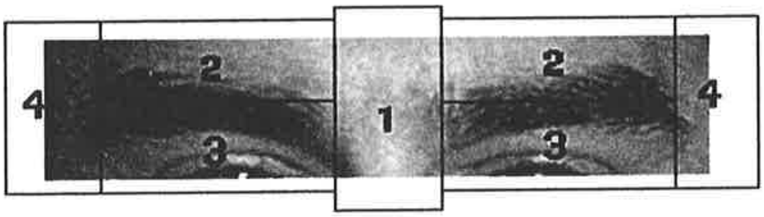


Figure 29: Illustration of the zones used to describe eyebrow plucking/waxing.

F-tests and histograms were used to compare data before the use of the relevant t-test (either equal or unequal variance). Significance was initially set at the 95% confidence level but altered according to the Bonferroni adjustment. Pearson's correlations were also calculated between the position of superciliare from the lateral iris and the other measurements.

Results

Overall, superciliare was found to be approximately 2.7 mm lateral to the lateral iris. However, values showed a large distribution resulting in high standard deviations (1 SD = 5.0 mm). Table 19 summarises the results for each sample. No statistically significant differences in horizontal superciliare location were detected between the sides or the populations of origin. There were, however, statistical significant differences between the sexes for all populations (P<0.017). On average, male superciliares were 3.5mm more lateral to the iris than in females and had smaller standard deviations (Table 19). Non-metrically, 80% of right superciliares could be found in the region between the pupil center and the exocanthion. For the left superciliare, almost 70% could be found between the pupil center and the exocanthion.

Table 19: Summary table of horizontal distance from superciliare to the lateral iris for males and females separated by population of origin.

		Right side (mm)		**Left side (mm)**	
	n	Average	S.D.	Average	S.D.
Male European	27	4.4	2.7	4.2	3.7
Female European	48	0.7	4.8	0.9	6.3
Male Asian	20	4.9	4.5	5.8	2.5
Female Asian	19	2.3	5.2	2.5	5.1
Other male individuals	7	6.4	3.6	4.1	3.0
Other female individuals	7	0.4	4.4	-0.8	2.8
Total Male	54	4.8	3.5	4.8	3.2
Total Female	74	1.1	4.9	1.2	5.8
Total	128	2.7	4.7	2.7	5.2

Other measures showed weak correlations with superciliare, however, actual superciliare values were disparate and therefore attempts at its prediction were unsuccessful. As a result, this aspect of the study will not be further discussed.

Overall, few males reported eyebrow manipulation. Of those that did (15% of male Europeans and 5% of Asians) most removed hair in zone 1 (see Fig. 29). Sixty-five percent of female Europeans removed eyebrow hair, 23% of this group removing hair in zone 2. Eighty four percent of female Asians removed eyebrow hair, 47% removing hair from zone 2. The female European sample was the only group large enough to test for differences between those that removed portions of their eyebrow and those that did not. No statistically significant results were found between these groups, although averages were closer to the lateral iris for those females who removed eyebrow hair (Right side = 0.13mm, Left side = 0.32mm) in comparison to those that did not (Right side = 1.79mm, Left side = 2.02mm). This suggests that these females may have represented the traditional facial approximation guideline well only because they plucked there eyebrow hair following cosmetic guidelines (e.g., Campsie, 1997; Campsie, 1998) which happen to be similar to the facial approximation guideline that has been suggested. Standard deviations were also slightly greater for the hair removal group (Right side = 4.88mm, Left side = 6.96mm) than the group that did not remove hair (Right side = 4.65mm, Left side = 4.84mm).

Discussion

The guideline that superciliare is located directly above the lateral iris may appear to be fairly accurate, especially in females, since small differences were found between the average horizontal position of superciliare and the lateral point of the iris (1mm for females and 5mm for males). However, the large variation in superciliare position, as reflected by large standard deviations (especially for females), and the observation that up to 30 percent of people's superciliares (about 1 in 3) fall outside of the region between their exocanthion and the pupil center demonstrate the large inaccuracy of this guideline and its limited usefulness. It seems that females only approximate the guideline closely because they pluck their eyebrow hair using cosmetic guidelines that are similar. It is suggested, therefore, that the guideline that the horizontal displacement of superciliare is equal to the lateral iris, is only used in a very general sense, if at all, in future facial approximation methods.

Since the lateral limbus of the sclera/cornea could not be precisely located from frontal photographs, the accuracy in using this landmark to predict the horizontal position of superciliare could not be evaluated. Since the limbus may fall slightly lateral to the lateral border of the iris, this guideline maybe slightly more accurate than using the lateral iris border itself. However, the magnitude of the measurements found for males in this study would suggest that this guideline is also inaccurate, at least for predicting male superciliares. Overall, it seems that this guideline should also only be used in a very general way to predict superciliare position and practitioners should note the wide range of variability in actual superciliare location.

It appears that the finding of a large proportion (up to 80% in this sample) of people's superciliares falling between the region of the exocanthion and the pupil center will be useful to facial approximation practitioners. The finding that superciliare is, on average, 2mm lateral to the lateral iris, and that this distance increases for males will probably also be useful.

It should be noted that some difficultly was encountered in the placing of superciliare due to the structure of the eyebrows, which are not generally well-defined arches. The individual hair fibers in the superior mid-portion of the eyebrow are angled inferiorly and laterally, becoming denser toward the center of the brow, but the density transition may be smooth or abrupt (Fig. 30). In smooth brows it is likely, from a distance, that the judgment of the superior part of the brow arch would be lower than it is in reality due to the scarcity of superior hair follicles. This placement of superciliare may be higher, from a distance, than the relative placement of a "false superciliare" in a person who has an abrupt density change (Fig. 30). Also, technically defining superciliare seems to be generally quite difficult since eyebrow hairs were observed to be present, although sparse, surprisingly high up on the forehead. Therefore, the determination of superciliare depends on the subjective interpretation of where hair density justifies it (Fig. 30 illustrates where superciliare was positioned in this study). The approach was taken in this study to define superciliare as being the highest point on the eyebrow that is readily determinable when the whole face was in view on the computer monitor (however, as seen in Fig. 30, numerous other hairs were present above and lateral to the position of superciliare).

Figure 30: Example of density change in male eyebrows
(a) smooth density transition to the "main brow" **(b)** abrupt density transition to the "main brow". Black arrow indicates region of superciliare. White arrow indicates region of the "false superciliare".

It is necessary to acknowledge that photogrammetric methods are limited by magnification and perspective distortions inherent to all photography methods (Farkas, 1994b; Farkas *et al.*, 1980; Gavan *et al.*, 1952; Iscan, 1993). However, measures of superciliare to the lateral iris border in this study are expected to be affected little since a large focal length lens was used, measures of superciliare and lateral iris borders were close to the camera focal plane at glabella, and the distance being measured was small (so absolute error in comparison to larger measures was much less). The validity of photogrammetric measures of the eyes is also supported by other studies that have found these measures not to differ from direct anthropometric measures in living subjects (Farkas, 1994b; Farkas *et al.*, 1980).

Although additional suggestions have been made for determining other features of the eyebrow it appears that these should be regarded with some caution until they have been tested and verified. These guidelines include: the medial most point of the brow falls vertically in line with the alare (Brennan, 1980; Ellenbogen, 1983); the lateral most point of the brow falls in line with a tangent connecting alare to the exocanthion (Brennan, 1980; Ellenbogen, 1983); the medial and lateral brow should fall horizontal to each other (Brennan, 1980; Ellenbogen, 1983); in males the brow arc is at the supra-orbital rim and in females it is above it (Brennan, 1980); individuals with strongly developed supraorbital margins have lower brows (Fedosyutkin and Nainys, 1993); strongly developed supra-orbital margins and brow ridges indicate an acute angle of the brow arch where less developed supra orbital margins and brow ridges indicate a more smoothly arched brow (Fedosyutkin and Nainys, 1993).

This sudy comfirms observations of Rosprym (1934) that eyebrow form is highly variable. However, since Rosprym's (1934) study was conducted over 75 years ago, current experiments that aim to test the repeatability of Rosprym's categories and character frequencies on modern samples would appear to be useful.

Average Face Morphology

Since facial approximation methods closely rely on averages for estimating facial features from the skull, like average soft tissue depths, the faces built will tend toward the average and therefore, should all appear somewhat similar, that is, average. However, facial approximations rarely appear to look similarly average. The lack of references in the facial approximation literature to articles describing average human faces (e.g., Alley and Cunningham, 1991; Langlois and Roggman, 1990; Langlois *et al.*, 1994; Little *et al.*, 2001;

Penton-Voak *et al.*, 1999; Perrett *et al.*, 1994; Rhodes *et al.*, 1999; Rhodes and Tremewan, 1996; Rowland and Perrett, 1995) suggests facial approximation practitioners are not generally aware of the anatomical appearance of average human facial features, and use their own subjective interpretations.

Additionally average human faces appear to be useful to facial approximation processes since they can be used to objectively create faces from skulls, particularly when using computers. Current computer generated facial approximation techniques generally involve 3D representation of the skull and soft tissue depths to which a facial surface is warped (Vanezis *et al.*, 2000). At present, the final approximations are surface colour free since for any given skull this information is not known (Vanezis *et al.*, 2000). It would be possible, however, to map the 3D facial surface with an average facial colouration appropriate to the individual's age and other characteristics. Although this would result in facial approximations that are somewhat similar looking, it would, at present, be the most objective way of creating facial approximations according to the American method that uses average soft tissue depths (Prag and Neave, 1997). The warping of average colour and texture information to 3D computer facial approximations may also increase their recognition accuracy compared to current computer techniques since slow-varying intensity patterns (colour and texture information) that are known to contribute to facial recognition (Bruce and Langton, 1994; O'Toole *et al.*, 1997) would be included.

Evenhouse and colleagues (1990) have attempted to use average faces in facial approximation methods before, but results have been less than optimal because nonstandardized face images were used, the samples used were extremely small (n=5), and averaging techniques they used were not entirely appropriate and have become outdated (e.g., average dimensions describe by Farkas (1981) were used to generate whole average faces). The aim of the studies following is to generate accurate standardized average faces of Australians by using highly standardized photographs and enhanced computer averaging techniques that can be used for modeling facial approximations.

Standardized Average Face Colour, Texture and Shape

Introduction

Facial averages can be generated by collecting two-dimensional (2D) photographs of faces and blending them together. Many 2D average faces have previously been generated, mostly for use in psychological studies (e.g., Alley and Cunningham, 1991; Kujawa and Strzalko, 1998; Langlois and Roggman, 1990; Langlois *et al.*, 1994; Penton-Voak *et al.*, 1999; Perrett *et al.*, 1998; Perrett *et al.*, 1994; Rhodes *et al.*, 1999; Rhodes and Tremewan, 1996) but there has been one attempt with regard to forensic facial approximation as described above (Evenhouse *et al.*, 1990). Facial averages have also been generated from 2D outlines (Rabey, 1977-78) and in three-dimensions using laser scanned images of faces (McCance *et al.*, 1997a; Tiddeman *et al.*, 2000; Tiddeman *et al.*, 1999).

The ability for facial averages to be representative of reality is dependent upon: (a) the quality of the original images/photographs (e.g., resolution, standardization, head position); (b) alignment methods used during averaging (e.g., rigid body registration and warping); and (c) blending process (e.g., texture preservation). Consequently, some averages may not be realistic since the methods used to collect and/or average the original images may not be appropriate.

Early attempts at generating average images consisted of superimposing facial images at various picture opacities (Grammer and Thornhill, 1994; Kujawa and Strzalko, 1998; Langlois and Roggman, 1990; Langlois *et al.*, 1994) (Fig. 31). However, this method did not generate average faces well, rather it displayed the variability within the face set with the more "solid" parts of the face representing standard deviations more than averages (Fig. 31). Also, the colour information displayed in these faces is not average since corresponding points on the faces may not be averaged together. For example, the colour information of the lips of one face image may be averaged with the colour information of part of the cheeks of another face image. Often this method also involves the manipulation of individual facial images to make the interpupillary distance and the height of stomion the same (Grammer and Thornhill, 1994; Langlois and Roggman, 1990) so that the clarity of the average image is higher (Fig. 31). However, this image manipulation changes the size, proportions and individual features of the face (see Fig. 31), and is valid only if they replicate the dimensions of the true average face (Alley and Cunningham, 1991; Rowland and Perrett, 1995).

(1) (2) (3) (4) (5) (6)

Figure 31: Examples of "Average faces" made from the same 32 female Europeans using different techniques
(1) straight superimposition aligning at nasion; **(2)** superimposition after normalizing faces based on pupil width (not average) and aligning faces on the pupils before blending; **(3)** superimposition after normalizing faces on pupil width and stomion position (not average) before aligning images based on these points and blending; **(4)** "Average face" resulting from methods similar to those used by Evenhouse and colleagues (1990) employing the "average face template" of Farkas (1981); **(5)** calculated average using a linear averaging method; **(6)** calculated average made using additional algorithms to preserve texture information.

Evenhouse and colleagues (1990) used average dimensions and templates describe by Farkas (1981) to generate whole average faces by warping and blending 5 faces to this template (see Fig. 31). However, Evenhouse and colleagues (1990) approach is less than optimal since average colour information was only generated for five faces which is probably not representative of population means. Also average feature shapes are not represented in Evenhouse and colleagues (1990) methods since Farkas did not measure them, rather he only measured facial proportions.

Since these earlier attempts method of generating average faces has improved. The approach taken now is to calculate the average face shape from x,y coordinates of individual images before warping the individual images to the average shape and blending the colour information together (Fig. 31). This method was pioneered by Perret and colleagues (Benson and Perrett, 1992; Penton-Voak *et al.*, 1999; Perrett *et al.*, 1998; Perrett *et al.*, 1994; Rowland and Perrett, 1995; Tiddeman *et al.*, 2001) and other authors have employed similar methods (Kujawa and Strzalko, 1998; Rhodes *et al.*, 1999; Rhodes and Tremewan, 1996). Here this method will be referred to as the calculated average technique. This method is advantaged because actual two dimensional face shapes are calculated and the colour information at corresponding points is averaged. It is worth noting that average faces have been produced using similar techniques in three-dimensions (Tiddeman *et al.*, 1999).

Image collection: Photographic techniques

To enable the generation of accurate, reliable and repeatable average faces the photography method used must be highly standardized. Differences in camera position, subject position, subject distance to the camera, type of lens, lighting and film development will alter photographs between sessions and will dramatically limit the usefulness of any comparisons (Dobrostanski and Owen, 1998).

To enable life-like images to be generated, a large focal length lens (e.g., 105mm) and a large subject to camera distance must be used to minimize distortions due to perspective, like object magnification and false edge representation (Gavan *et al.*, 1952) (see p. 41 of this thesis). It has been suggested that when using a 100mm lens, a distance of approximately 1.5m gives an image similar to that seen during normal social contact (Dobrostanski and Owen, 1998).

Use of cameras with small focal length lenses (e.g., <100mm), or small subject-camera distances (e.g., <1m) will result in images that are not well balanced. For example, features of the face that are closer to the camera (like the nose) will appear prominent and unrealistically large. In addition the true edges of facial features (like noses, head outline, and ears) will not be seen. Furthermore, features further from the camera, like the ears, appear proportionately and unrealistically small as previously described (Fig. 5).

Few previously published averaging papers adequately describe the photographic methods they use making it difficult to evaluate what the "standardized procedures" apparently employed were and how realistic the final images actually are. For example, many papers simply state that frontal photographs, with neutral expression, were digitized (e.g., Langlois *et al.*, 1994; Perrett *et al.*, 1994; Rhodes and Tremewan, 1996). Others go as far as saying that faces were also photographed under the same/standard lighting conditions (Kujawa and Strzalko, 1998; Perrett *et al.*, 1998; Rowland and Perrett, 1995) and at standard distances (Langlois and Roggman, 1990), but do not specify what the standard conditions actually were. This makes it difficult to compare these studies and/or repeat their results. Also, even if the conditions used are reported as being standardized, they may not be appropriate for realistic facial recordings, but one cannot tell since the exact conditions are not reported.

Although it has been suggested that standardization for lighting, view, facial expression, and makeup, is not necessary when using samples larger than 30 individuals (Rowland and Perrett, 1995), this does not argue against using standardized images. While it is true that a small number of images that deviate from standardized conditions have little power to affect averages made from large numbers of highly standardized photographs, the case seems to be different for small samples of partially standardized images. Partially standardized photographs are likely to include variations between many images increasing their power to affect the mean. Such variations are unlikely to be random (due to photographers technique etc.), particularly in smaller samples (n<50), resulting in different means in comparison to highly standardized images. Consequently, images should be as highly standardized as possible.

Image collection: Head position

Not only do photographic techniques need to be appropriate and standardized to generate 2D average faces that are representative of reality, but the position of the head must also be standardized. If facial anatomy is to be recorded as other individuals normally see it, photographs must be taken with the head in a position commonly adopted by individuals. Therefore, the natural head position or the position of the head "when a person is standing with his visual axis horizontal" (Broca, 1862) should be used since individuals are represented as they naturally appear in life (Cooke and Wei, 1988b; Moorees and Kean, 1958; Moorees *et al.*, 1976).

There are many other reference lines that can be used to standardize head position, the Frankfurt horizontal being one that has previously been used for generating facial averages (McCance *et al.*, 1997b). Others (Rabey, 1977-78; Tiddeman *et al.*, 1999) have also used a reference line that is comparable to the Frankfurt Horizontal. However, these reference lines are not optimal since they do not represent the typical position in which the head is normally held, and they assume stability of reference points that are, in reality, subject to biological variability (Garn, 1961; Moorees *et al.*, 1976), e.g., in the Frankfurt horizontal three reference points are used, both porions and left orbitale. The Frankfurt horizontal is additionally limited because it assumes that the transmeatal axis is perpendicular to the midsagittal plane (Cavallaro *et al.*, 1974). The use of the Frankfort horizontal in living subjects also appears to be somewhat illogical since it was derived for the purpose of orienting skeletal remains that could no longer assume an upright posture (Moorees *et al.*, 1976) and in the living a natural head posture can be directly obtained (Moorees *et al.*, 1976).

The natural head position also has the advantage to other planes because it is more repeatable (even after 15 years (Peng, 1999)) having an error of about 2 degrees (Cooke, 1990; Peng, 1999; Siersbaek-Nielsen and Solow, 1982) in comparison to the Frankfurt horizontal and the nasion-sella line, which have an error of about 5 degrees (Cooke and Wei, 1988a; Cooke and Wei, 1988b; Moorees and Kean, 1958; Solow and Tallgren, 1971).

Image alignment

The lack of highly standardized photography methods seems to have lead to photographs being "normalized" to compensate for any differences in size and to control for orientation. Often interpupillary distance is used for image alignment (Langlois and Roggman, 1990; Penton-Voak *et al.*, 2001; Perrett *et al.*, 1994; Rowland and Perrett, 1995), and a midpoint along a tangent connecting the pupils has also been used (Penton-Voak and Perrett, 2000). Unless the average interpupillary distance is calculated prior to and used for normalization, the size of the final average image will not be representative of reality (Rowland and Perrett, 1995).

Normalization techniques, like reference lines for determining head position, assume that the registration points used (often the pupil centers) are stationary, or fixed, across faces, which is not the case (Garn, 1961; Moorees *et al.*, 1976). Since registration techniques require assumptions to be made that are not generally true across individuals, it is best to use as few registration points as possible to reduce the number of assumptions made, e.g., a single midsagittal point is preferable to two or three bilateral points.

When photographs are standardized with all individuals identically positioned about a single midsagittal point (e.g., glabella), with two direction vectors (e.g., vector of the individual facing the camera and the downwards vector of gravity), no post-photographic image registration/normalization is needed. This is advantageous because it limits the number of assumptions made regarding the fixation and symmetry of reference points. Although it appears to be possible to align images using a midsagittal point after photography, in practice, it is difficult since most midsagittal anthropometric points cannot be accurately determined from frontal or profile images alone (Farkas, 1994b; Farkas *et al.*, 1980).

Blending processes

The realistic appearance of the anatomy of facial averages has also previously been limited by the blending process itself since textural information is lost during image blending, resulting in unrealistically smooth images (Tiddeman *et al.*, 2001). However, this can be reduced using additional algorithms (Tiddeman *et al.*, 2001).

The aim of this study was to generate high quality two-dimensional images of average human face morphology of Australians from Melbourne that are closely representative of reality by using strictly standardized craniofacial photography, natural head position, and averaging techniques that retain texture information, but without the use of any post-photographic normalisation methods.

Averages were generated for Australian individuals grouped into 4 categories according to self-perceived ancestry: (a) Male Asians; (b) Female Asians; (c) Male Europeans (d) Female Europeans. The averages are not intended to be standards of Asian or European face morphology but an illustration of socially perceived, population-specific morphotypes.

Methods

Sample

The sample consists of 152 individuals, aged between 18 and 34 years, living in Melbourne Australia. All gave informed consent before taking part in this study. The participants were grouped according to their perceived main ancestral background and sex. Seven individuals reported mixed ancestries and so were not included in the averaging process, giving a sample size of 145 individuals.

Four groups were established to classify the sample according to self-reported population ancestry: Female European (n=57 at rest, n=56 smiling), Male European (n=29 at rest, n=27 smiling), Female Central/South East Asian (n=28 at rest, n=27 smiling) and Male Central/South East Asian (n=31 at rest, n=30 smiling). Some sample numbers differ between smiling and non-smiling scenarios since several subjects failed to smile and/or stay correctly positioned during photography of the smiling pose.

Photography

Participants were photographed, in the natural head position, in smiling and neutral expressions, in both frontal and right side profile, on the craniofacial photography rig described above (p. 39 to 62).

Although participants were photographed individually they were brought into the photography room in pairs. The presence of a friend seemed to make smiling more natural to the subject being photographed. However, care needed to be taken to ensure subjects were correctly positioned since they would often lean towards the camera when laughing or smiling (one reason for smaller smiling sample sizes in this paper). The advantage of a friend being present was that a natural smiling expression seemed to be obtained rather than a "posed" smile.

Film development was standardized as much as possible by having the same qualified photographer develop the photographs in house. Once the slides were processed they were scanned into a computer using a Nikon® SF-2000 slide scanner. Several preferences in Nikon® Scan were selected (e.g., clean image function was set to normal, bit depth was set to 8, multisample was set to 4x (fine), interpolation was default, colour space was RBG Adobe 1998, and autofocus and autoexposure were turned on) during image scanning. These preferences were held constant during the scanning of the slides. The resultant pictures were 1,200 pixels in width, 1,803 pixels in height and were originally saved in TIFF format before being converted to JPEG format. JPEG conversion was used for easier file management. Since the images were very large natural images, conversion to JPEG format did not appear to affect their visual appearance.

Delineation

The faces were scanned into a computer and were then delineated, by hand, before averaging took place. Both the delineation and averaging were done using "Psychomorph" software developed by Perrett et al. (Benson and Perrett, 1992; Rowland and Perrett, 1995; Tiddeman *et al.*, 2001). The delineation process involved placing landmarks at certain locations on the face, some being standard anthropometric points (Fig. 32 and 33, and Appendices 6 and 7). Many of the landmarks were joined together by contour lines that gave the outline shape of the face (Fig. 32 and 33). These outlines will be referred to as "delineation maps" of the face. Although automated delineation methods are possible (Vetter *et al.*, 1997), they are not always exact, and therefore a manual approach was taken here to ensure anthropometric landmarks were correctly placed at specific anatomical locations.

The delineation maps were originally adapted from Brennan (Rowland and Perrett, 1995). In the past these delineation maps, have included 174 landmarks in frontal view. For this study, we modified an existing frontal delineation map and created a new profile map. The new frontal delineation map included 219 points (Fig. 32) placed at strategic anatomical locations (Appendix 6). The inclusion of more landmarks in the new delineation map enabled the shape of many more morphological features to be included. The new profile delineation template included 147 landmarks (Fig. 33) and outlines similar features as described by frontal delineation maps (Appendix 7).

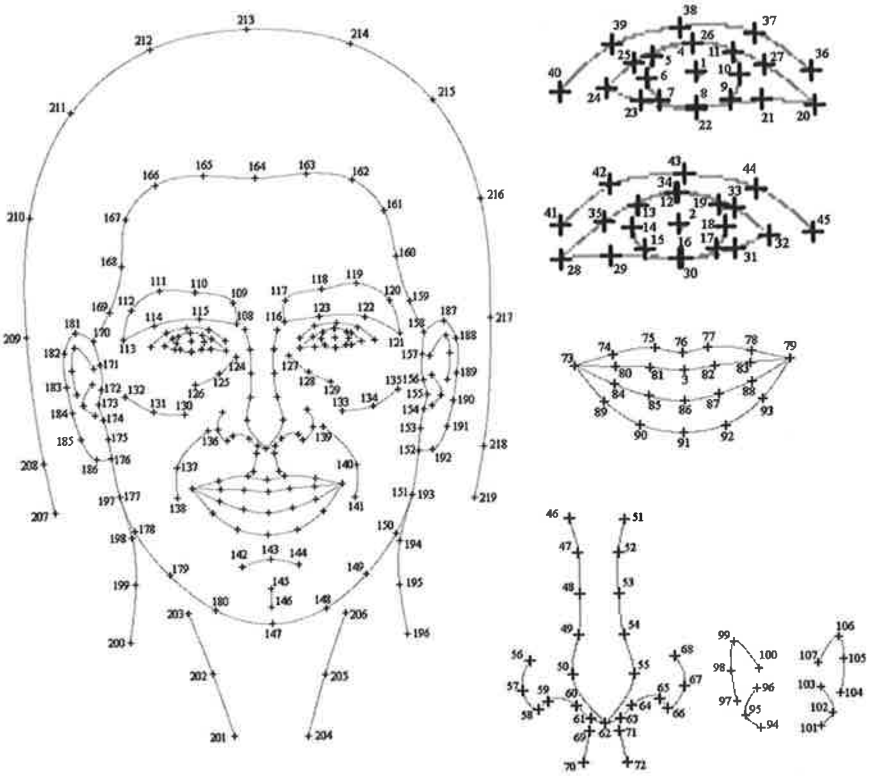


Figure 32: Frontal delineation map including 219 points (see Appendix 7 for landmark descriptions)

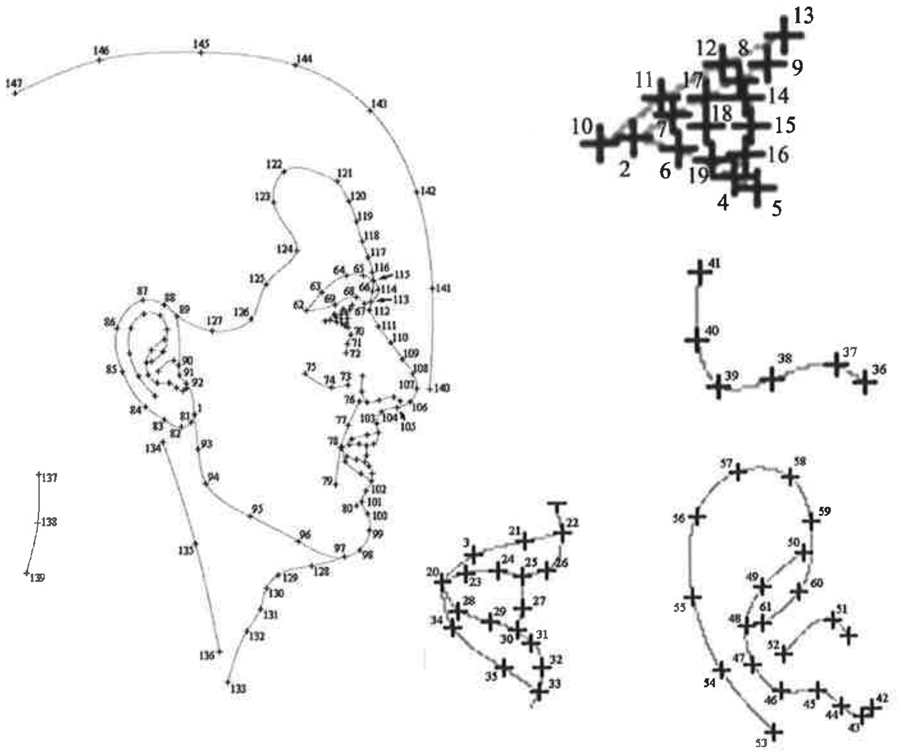


Figure 33: Profile delineation map including 147 points (see Appendix 8 for landmark descriptions)

Averaging

The averaging procedure had several steps. First the computer generated an average delineation map from the individual maps by calculating the average x and y coordinate for every point making up contour lines (Fig. 34). No normalization was required in this step since all faces had been photographed under repeatable conditions using glabella for alignment. The computer warped each individuals face into the average template using a multi-scale algorithm (Tiddeman, 1998) (Fig. 34). The average colour (with red, green and blue components) of each pixel was calculated to produce the initial average face image (Fig. 34). To preserve textural detail in the blends an additional algorithm was applied to amplify the edges (at different positions, spatial scales and orientations) to the appropriate amount for the sample (Tiddeman *et al.*, 2001) (Fig. 34). The resultant average faces had the same resolution as the original input faces, being 1200x1803 pixels.

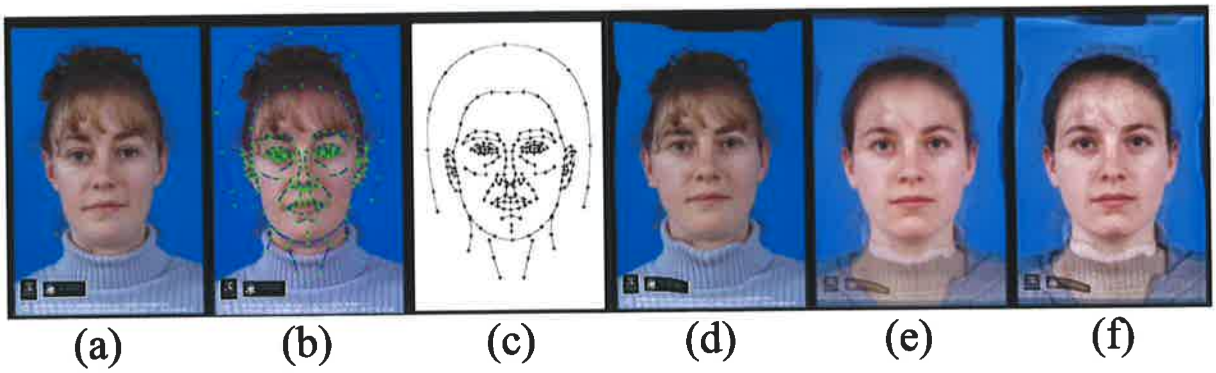


Figure 34: Examples of face averaging
A standardized face photograph **(a)** is delineated **(b)**. The face delineations for each photograph in the sample is averaged **(c)** and individuals' faces are warped to the average shape **(d)**. Faces warped to the average face shape are then blended together, **(e)** and **(f)**. **(e)** – shows the blend of face **(a)** with two other faces without texture preservation and **(f)** – shows the same blend with texture preservation.

Since direct anthropometric measures do not exist for the sample studied, measurements made from the final average images were compared to direct anthropometric measures made by Farkas et al. (1994a; 1994b) on similar ancestral groups but from North America and Singapore, to check for accuracy. Since it has been found that only eye and mouth measures made photogrammetricly compare to direct anthropometric measures (Farkas, 1994b; Farkas *et al.*, 1980) we expected only these to be similar if the average faces were to be representative of reality, however we made many other measures also (see Table 20).

Measurements taken from the scanned photographic images (mm) were scaled to life size by a factor of 4.975124 as determined above and were also adjusted by 1% for every 10mm the feature fell behind or in front of the camera focal plane as determined above. Since profile and frontal images were taken simultaneously feature distances from the focal plane could be determined and therefore feature measurements could be roughly corrected for scale according to the above information.

Results

Figure 35 shows the frontal averages generated for each population of origin. Figure 36 shows the average faces for each population of origin in profile view. Figure 37 shows enlarged facial features that demonstrate the high resolution of these pictures. These pictures represent average human anatomy as normally seen from the frontal and profile views because the photographs were highly standardized, participants were photographed in the natural head position, post-photographic normalization was not used in the averaging process, and texture detail was retained. The high quality of these average images is evident from the photographs since single flash unit reflections can be seen on the eyes, and facial characteristics are distinct, e.g., medial canthal ligaments can easily be seen, and even individual hairs in the eyelashes and eyebrows are evident. The successful generation of such high quality images has resulted from the high level of photographic standardization, combined with the use of improved computer graphic methods.

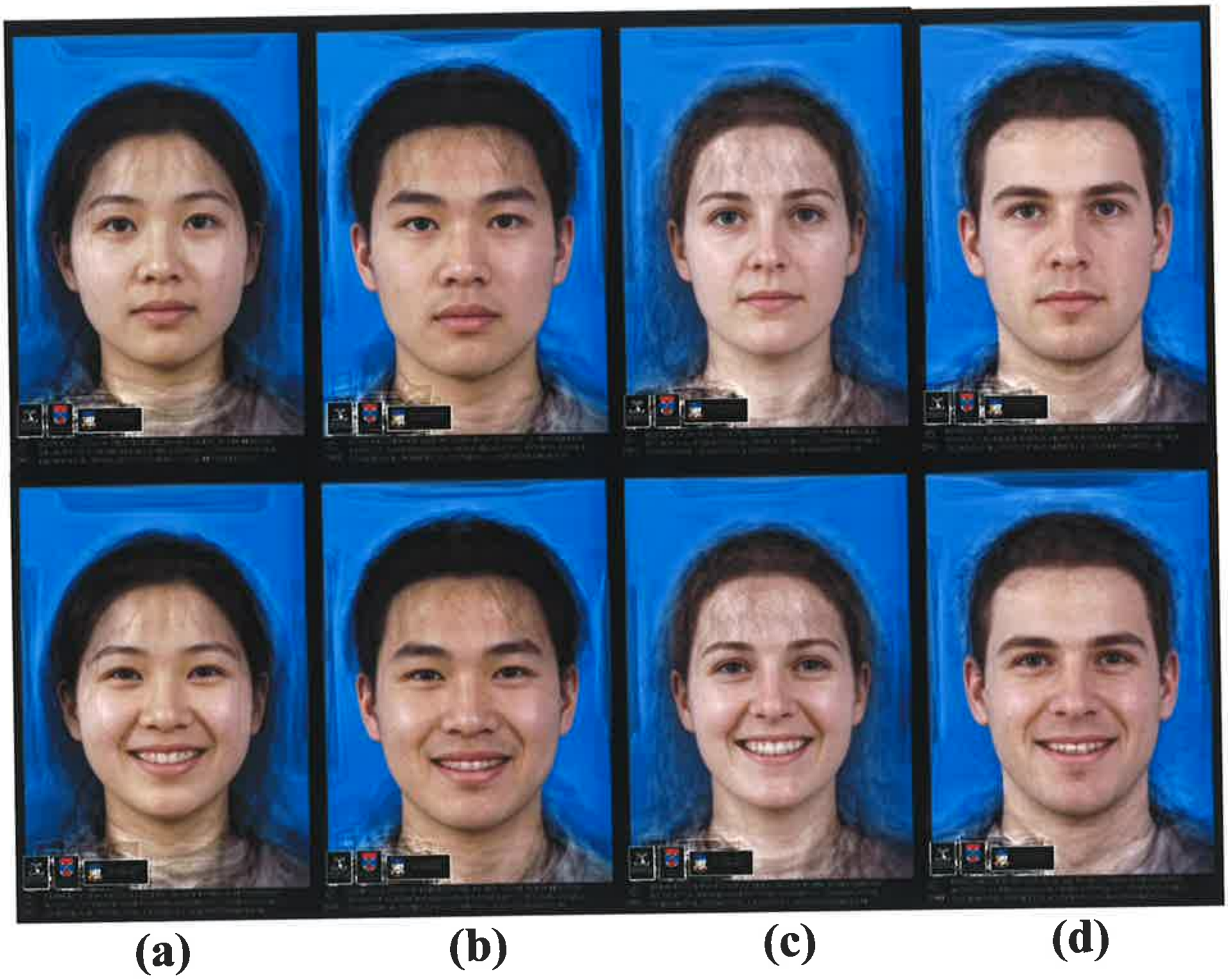


Figure 35: Average frontal faces generated for each group
Top row shows the average with neutral expression, lower row shows smiling expression. **(a)** female Central/South East Asian: Top n=28 (aged 20 ± 2 years); bottom n=27. **(b)** male Central/South East Asian Top n= 31 (aged 20 ± 3 years); bottom n=30. **(c)** female European Top n=57 (aged 23 ± 4 years); bottom n=56. **(d)** male European Top n=29 (aged 22 ± 4 years); bottom n=27.

(a) (b) (c) (d)

Figure 36: Average profile faces generated for each group
Top row shows the average with neutral expression, lower row shows smiling expression. **(a)** female Central/South East Asian: Top n=28 (aged 20 ± 2 years); bottom n=27. **(b)** male Central/South East Asian Top n= 31 (aged 20 ± 3 years); bottom n=30. **(c)** female European Top n=57 (aged 23 ± 4 years); bottom n=56. **(d)** male European Top n=29 (aged 22 ± 4 years); bottom n=27.

(a) (b) (c) (d)

Figure 37: Enlarged average face features to show detail
(a) enlarged eye (female European) showing flash reflection, medial canthal ligament and maybe even individual eyelashes. **(b)** enlarged eyebrow (male Asian) showing well defined hairs. **(c)** enlarged nasolabial fold (female European). **(d)** enlarged tip of nose (male European) showing skin pigmentation possibly caused by pore locations.

Table 20 presents metric measures taken on each of the average faces (adjusted for scale and magnification) and their comparisons to direct anthropometric measures taken by Farkas *et al.* (1994a; 1994b) on other European populations. It can be seen that measures not only of the eyes and mouth compare closely to the 95% confidence range of the population mean taken by Farkas *et al.* (1994a; 1994b) but also measures of other features like the ears compare closely. This indicates these average images (and the photographs from which the averages were generated) are representative of reality.

Discussion

The average images presented here are useful reference guides for facial approximation because they provide objective average face information, which previously has been determined subjectively, particularly for features not yet known to be related to the skull. It also appears that because soft tissue facial morphologies are unlikely to ever be exactly estimated from the skull, facial approximation practitioners will be forced to rely on such averages despite their disadvantages, i.e., (i) averages are not representative of unique individual appearances (Brues, 1958; von Eggeling, 1913), (ii) prediction error is increased for some individuals when single population averages are used for trait/s that are not normally distributed, e.g., display binomial distributions, (iii) averages increase similarity between facial approximations and, therefore, probably make specific recognition more difficult, (iv) the application of average methods to less typical, or more distinct, skulls decreases the distinctiveness of the face and makes those faces harder to recognize (Bruce and Young, 1998).

Table 20: Comparison of average face measures (adjusted) to those taken on similar populations by Farkas *et al.* (1994a; 1994b)
Ninety-five percent confidence intervals of Farkas et al. of population means are represented by the 1.96 SEM; en-en to sa-sba (l) are taken from frontal images; n-gn to sn-sto are taken from profile images.

Measurements	**Male European**			**Female European**			**Male Asian**			**Female Asian**		
Frontal	Farkas et al.	1.96 SEM	Av. Face	Farkas et al.	1.96 SEM	Av. Face	Farkas et al.	1.96 SEM	Av. Face	Farkas et al.	1.96 SEM	Av. Face
en-en	33.3	0.5	33.1	31.8	0.3	31.0	37.6	1.2	37.0	36.5	1.1	35.8
ex-ex	91.2	0.6	88.5	87.8	0.4	85.8	91.7	1.4	92.2	87.3	1.9	86.8
en-ex	31.3	0.2	28.0	30.7	0.2	27.9	29.4	0.5	27.9	28.5	0.6	26.2
ps-pi	10.8	0.2	11.1	10.9	0.2	11.1	9.4	3.4	10.0	9.5	0.4	11.0
al-al	34.9	0.4	37.8	31.4	0.3	33.8	39.2	1.0	41.8	37.2	0.8	38.3
ch-ch	54.5	0.6	51.7	50.2	0.5	48.8	48.3	2.4	49.5	47.3	1.2	48.0
ls-sto	8.0	0.3	6.0	8.7	0.2	5.5	11.2	0.4	8.4	10.1	0.5	7.4
sto-li	9.3	0.3	9.4	9.4	0.2	8.5	10.8	0.5	10.9	10.5	0.5	10.4
sa-sba [r]	62.7	0.7	61.4	59.6	0.5	56.1	61.0	1.3	61.4	58.8	1.3	54.3
sa-sba [l]	62.9	0.7	62.5	59.9	0.5	56.7	60.7	1.4	63.0	57.6	1.4	57.5
Profile												
n-gn	124.7	1.1	125.7	111.4	0.7	116.4	123.6	1.9	124.0	114.9	1.8	113.4
n-sto	76.6	0.8	75.6	69.4	0.4	70.2	78.2	1.4	75.0	71.8	2.0	69.7
sn-gn	72.6	0.8	72.0	64.3	0.6	66.5	72.7	1.9	72.4	66.4	2.0	65.0
sto-gn	50.7	0.8	50.1	43.4	0.4	46.3	53.4	1.5	49.0	47.2	1.2	43.7
prn-gn	91.7	1.1	85.0	81.4	0.6	78.0	88.8	1.8	84.4	81.2	1.5	76.4
g-sn	67.2	0.9	68.8	63.1	0.6	68.1	66.5	1.3	74.0	62.3	1.6	71.8
t-ex	85.3	0.6	78.0	78.9	0.5	76.5	87.3	1.4	81.9	82.5	1.1	77.4
n-sn	54.8	0.6	53.7	50.6	0.4	50.4	53.5	1.0	51.6	51.7	1.2	48.4
sn-sto	22.3	0.4	21.9	20.1	0.3	19.8	23.5	0.8	23.4	21.6	0.8	21.3

The average images generated in this study are likely to be more representative of real average facial appearances than other average faces presented elsewhere (e.g., Alley and Cunningham, 1991; Langlois and Roggman, 1990; Langlois *et al.*, 1994; Little *et al.*, 2001; Penton-Voak *et al.*, 1999; Perrett *et al.*, 1994; Rhodes *et al.*, 1999; Rhodes and Tremewan, 1996; Rowland and Perrett, 1995) because highly standardized craniofacial photographs, natural head position, and techniques that retain texture information were used without any post-photographic normalization methods. The accuracy of these images is indicated by the similarity between measures from the average face images and direct anthropometric measures (Table 20).

It is worth noting, that any abnormal rotation of subjects' heads away from the frontal plane, in the natural head position, (except for rotations within the focal plane of the camera, e.g., the coronal plane if images are taken with a frontal camera) would result in smaller average head sizes than in reality. Since participants could see their faces in a mirror when adopting their natural head position it seems likely that normal participant head positions were recorded and, therefore, that the average faces generated are representative of the average as would normally be seen in life (i.e., in the natural head position).

Generation of average faces in 3D, using highly controlled conditions and the natural head position would be a vast improvement on the current study since facial measurements from the average 3D face would be directly comparable to anthropometric measures giving a better indication of the averages accuracy. Additionally the inclusion of more individuals into the average would be an advantage as it would give a better representation of general human anatomy. Because of the limited sample sizes used in this study, the average faces presented here are expected to be most representative of people from Melbourne Australia and are probably not representative of people from disparate geographic locations. This indicates that other average faces should be produced for people from disparate regions and that the faces produced here should be used in the meantime.

It is worth noting that features evident in these average faces reflect findings of soft tissue studies reported above, e.g., that the width of the mouth is less than the width of the pupils and the position of superciliare is slightly lateral to the lateral iris, especially in males. These images are therefore useful because they allow these results to be visualized with respect to other average features rather than just appearing as numrical values.

It would not be surprising if readers find the average faces presented here attractive and also familiar. These are common findings of many other studies examining average faces (Alley and Cunningham, 1991; Langlois and Roggman, 1990; Perrett *et al.*, 1994; Rhodes *et al.*, 1999; Rhodes and Tremewan, 1996). It is also readily observable that the average faces are much more symmetrical than most individuals' faces due to the averaging of features where left and right differences are normally distributed about a mean of zero.

It can be seen from Figures 35 and 36 that the teeth are not clearly defined in the average images of the smiling faces. This is the result of the warping individual faces to the average shape without delineation of the teeth. The hair also appears to have a wavy appearance and again, this is due to a lack of delineation and the averaging process. These artifacts may be reduced by inclusion of more delineation points, however, these features are extremely varied and therefore not easily delineated, e.g., not all persons smile showing the same number of teeth, nor do they all have similar haircuts. These artifacts need to be addressed in future work so current techniques can be further improved.

It is also worth noting that the average faces made here do not indicate differences between "races" nor are they standards of European and Asian facial anatomy since the faces were not segregated into groups based on *real* genetic ancestry. Groupings were made based on subjects' *self-perceived* ancestry, excluding those who had mixed perceived ancestry. For this reason, the faces represent socially perceived facial morphotypes and are probably indicative of population stereotypes. It is possible that in reality some of the individuals are grouped incorrectly because their true genetic ancestry differs to their perceived ancestry. Despite this, categorizations based on self-perceptions are useful because actual genetic ancestries are rarely known for sure. Additionally these average faces indicate facial anatomy that is typical of the various socially perceived "population groups". Average faces generated according to such groupings, may also be useful to police and medical surgeons in addition to facial approximation practitioners, since the faces display objective information about stereotypical population groups that is probably more accurate than an individuals' subjective interpretation.

One limitation to the above averages is that it is unknown how many faces need to be included in the sample before a robust average can be obtained. This is the topic of the next chapter.

Sample Sizes Needed to Generate a Robust Face Average

Introduction

While many attempts have been made in the past to create images of the average human faces (see above) no one has attempted to determine what sample sizes must be used. For example, it is unknown if a sample of 30 male, European, University of Adelaide students, aged 18 years, is adequate to generate an average face displaying the true average facial anatomy of male, European, University of Adelaide students, aged 18 years. Furthermore, average human faces from small samples are unlikely to be representative of broader population groups, e.g., male Europeans from a particular city are unlikely to represent male Europeans country wide, nor may they be likely to represent male Europeans from other countries. Therefore, when specific samples are used to make averages, the averages should be specifically referred to, rather than being described just as "average human faces", which appears to be misleading. It certainly seems that Evenhouse et al. (1990) attempt by using 5 faces in an application for facial approximation is less than adequate for generating a generalized average face.

It is necessary to know the minimum number of individuals needed to generate reliable averages for particular groups of individuals if studies are to be repeatable and so the generality of experimental results, across broader groups, can be estimated. The samples previously used for averaging have tended to be small (n<50) and specific - often 1st year students from particular universities. Sample sizes have ranged from less than 20 (Little *et al.*, 2001; Penton-Voak *et al.*, 1999) to little more than 30 in many cases (Langlois and Roggman, 1990; Langlois *et al.*, 1994; Perrett *et al.*, 1998; Rhodes and Tremewan, 1996). The highest number of individuals used for generating an average face is reported by Perrett *et al.* (1994) who, in one case, used a sample size of 342 individuals.

While it would be best to determine how many faces are needed for the reliable representation of broad population groups, e.g., people of socially self-perceived European origin, this goal is unrealistic given the large sample that would be required and the time restraints of this project. Therefore, this study aims to establish what sample sizes are needed to create a reliable facial average of a very specific population sub-group: female European students at the University of Melbourne who are predominantly studying dentistry. This information is useful because it indicates the minimal sample sizes that must be used for broader population groups and gives some indication of how reliable average faces generated from such broader population groups are.

Methods

Average Face Stimuli

One hundred and thirty female volunteers, between the ages of 18 and 34 (mean age 22.3 years, sd 4.0 years), were photographed under highly standardized conditions using the craniofacial photography rig described above. The subjects were predominantly "white" dental students at the University of Melbourne.

The photographs were sorted into independent samples by randomly selecting individuals at each age year. Five different stimulus groups were created, i.e., samples consisting of 10 individuals, 26 individuals, 44 individuals, 60 individuals or 130 individuals, as described in Table 21. Faces in each sample were averaged using software developed by Perrett et al. (Perrett *et al.*, 1998; Rowland and Perrett, 1995; Tiddeman *et al.*, 2001) at the Perception laboratory, The University of St. Andrews. This resulted in 12 average faces in total: 3 averages made from independent groups of 10 individuals; 3 averages made from independent groups of 26 individuals; 3 averages made from independent groups of 44 individuals; 2 averages made from independent groups of 65 individuals; and 1 average made from a group of 130 individuals.

Table 21: Summary table of the average faces made in this study

Average face no.	No. of people included in the averages	Average face code
1	10	10_1
2	10	10_2
3	10	10_3
4	26	26_1
5	26	26_2
6	26	26_3
7	44	44_1
8	44	44_2
9	44	44_3
10	65	65_1
11	65	65_2
12	130	130_1

Average faces within each stimulus group were arranged so that they could be compared in pairs, side-by-side (Fig. 38). Facial images were printed on Epson photo quality ink jet paper (matt) using an Epson® Stylus 740 colour printer. Average face images (including the image border) were 200mm in height and 133mm in width.

(1) 10_1/2 (2) 26_1/2 (3) 44_1/2 (4) 65_1/2

(1) 10_1/10_1 (2) 44_3/130_1 (3)130_1/65_1

Figure 38: Average faces made from different sized samples

A. Examples of comparisons made between average faces made from independent samples. **(1)** example averages made from independent samples of 10 individuals. **(2)** example averages made from independent samples of 26 individuals. **(3)** example averages made from independent samples of 44 individuals. **(4)** example averages made from independent samples of 65 individuals.

B. Examples of comparisons between control images **(1)** and between the 130 person average and 44 **(2)** and 65 **(3)** person averages.

Study Protocol

To determine if the faces generated from independent samples of the same number of individuals were the same or not, face perceptions were statistically tested, rather than using multivariate analysis of face metrics. This was done because face recognition primarily depends on people's perceptions not necessarily exact anatomical similarities or differences in faces. However, some facial dimensions, like distance between chelions, exocanthions, and alares were also taken to determine if any differences existed.

Twenty trials (10 frontal and 10 profile views) were presented in a random order to 50 assessors (22 males and 28 females, mean age 32 years, SD 15.7 years) who were asked to indicate, in a forced choice scenario, if they thought the faces they were presented with (which were in fact averages made from independent samples of the same number of people) were of the same person or of different people. Trials consisted of combinations of average faces made from 10 to 65 individuals. Participants were unaware that the faces had been standardized and averaged. To ensure that facial recognition of participants was tested, and not their ability to determine if the images were identical, participants were urged to look at the overall facial appearance. It was also explained that the pictures may have been taken on different days and, therefore, some features may vary between the photos even if they were of the same person. Thirty-three participants (16 males and 17 Females, mean age 33 years, sd 13.5 years) were also tested with control images. These face comparisons consisted of two images of the same face but with one image being 5% smaller than the other. Testing scenarios included both frontal and profile views.

Recognition results were compared against expected chance levels for guessing (50%) and the recognition rates of the control images (truly same face images) using Fisher's exact tests included in the JMP® (3.0.1) statistical package. The average faces made from the smallest samples, that were perceived to belong to the same individual at rates statistically above chance and not statistically different to control rates would suggest the minimum number of individuals needed to make a robust average face. It may be expected that the recognition rates of control images would be around 70% since this is the rate recorded for correctly matching individuals in video footage to their respective facial photograph in face pools (Hancock *et al.*, 2000).

Results

It can be seen from figure 39 that when assessors were presented with the control images (two photos of the same exact person, but with one made 5% smaller) only 73% believed the frontal images to be of the same individual (statistically above chance rates $p<0.10$), and 67% of assessors perceived the profile images to be of the same individual ($p>0.10$). Since recognition rates were not 100%, it appears that in this study some subjects gave priority to pictorial rather than structural codes, as described by Bruce and Young (1986). These recognition rates are also similar to those obtained in other studies (70%) using an unfamiliar identification scenario of real people from video images to photographic face pools (Hancock *et al.*, 2000).

Averages made from 44 people were perceived to be of the same person, between 68% and 76% of the time for frontal views. These rates were statistically above chance in two out of the three scenarios ($p<0.05$, Fig. 39) and were not statistically different from control images in any case ($p>0.79$). In profile views, recognition rates decreased. Averages made from 44 individuals were only recognized as belonging to the same person 58% and 68% of the time, and while these rates were not statistically different to those of chance ($p>0.10$, Fig. 39) they also did not differ at statistically significant levels from the control images ($p>0.49$).

When the average faces were made from 65 individuals the recognition rate increased to 88% in frontal views, which was statistically above chance rates ($p<0.01$) and those of control images ($p<0.10$, Fig. 39). In profile views recognition rates increased to 70%, which was statistically above chance rates ($p<0.10$, Fig. 39) and not significantly different from rates of the control images.

Average faces made from 26 individuals were not recognized as the same individuals above chance levels in either frontal or profile views ($p>0.10$, Fig. 39). However, two frontal image comparisons were not recognized significantly less than control images ($p>0.05$) while one case (26_1/3) was ($p<0.01$). In profile there were two instances where recognition rates were below control rates at statistically significant levels ($p<0.05$) and one instance were it was not. Faces made from 10 individuals, in both frontal and profile views, were perceived as *not* belonging to the same individual above statistically significant levels when compared to chance ($p<0.05$, Fig. 39), or control images ($p<0.001$). Overall, profile views of average faces were consistently reported to be of different people at higher rates than frontal images ($p<0.05$). Even though averages made from small samples were almost always identified as not being the same individual, differences in metric dimensions between these averages were small (Table 22).

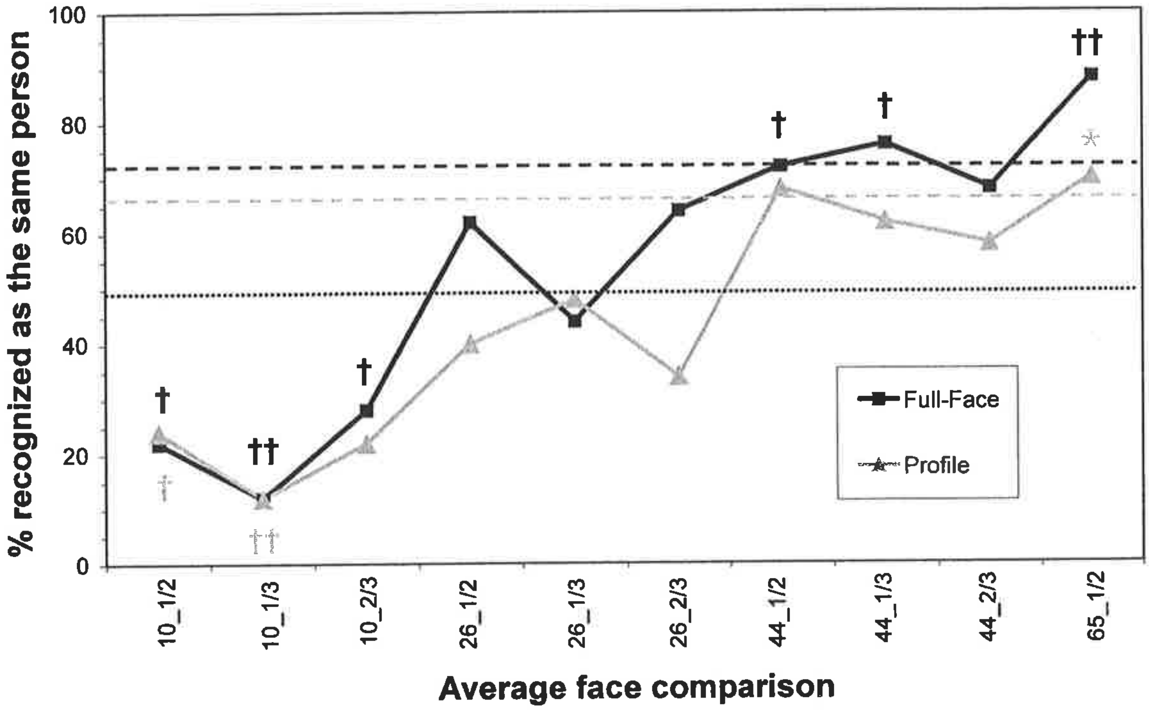


Figure 39: Frequency of average facial images being recognized as the same individual.

------	Chance rate
— —	Control recognition rate frontal view
- - -	Control recognition rate profile view
*	statistical significance at $p<0.10$ in comparison to chance
†	statistical significance at $p<0.05$ in comparison to chance
††	statistical significance at $p<0.01$ in comparison to chance

Table 22: Metric comparisons between the average faces

Av. Face	en-en	ex-ex	al-al	ch-ch	li-ls	sa-sa	sba-sba	sto-gn	sto-sn	sn-g
10_1	29.5	84.0	33.8	49.1	13.9	183.6	147.1	45.0	19.9	63.0
10_2	29.0	85.2	32.8	48.5	13.9	181.6	142.5	42.3	18.9	63.4
10_3	29.1	86.4	32.3	49.8	14.9	185.2	147.6	43.0	19.4	65.6
26_1	28.5	83.6	31.8	47.6	12.9	183.2	145.1	44.0	18.9	64.0
26_2	28.5	84.1	31.8	48.2	13.9	183.0	147.3	44.0	20.4	63.5
26_3	29.1	85.4	32.8	48.7	13.9	185.2	147.8	46.1	20.9	64.6
44_1	28.0	84.1	32.3	49.2	14.4	185.5	148.8	44.5	20.4	64.0
44_2	29.0	85.2	32.3	47.2	13.9	183.8	147.2	44.1	19.9	64.0
44_3	29.6	84.7	32.3	47.6	14.9	184.9	147.9	44.5	21.4	63.9
65_1	28.0	85.2	32.3	48.7	13.4	185.2	147.9	45.1	19.9	64.0
65_2	29.0	84.8	32.3	48.1	14.4	184.0	147.2	43.5	20.9	64.5
Average	28.9	84.8	32.4	48.4	14.1	184.1	146.9	44.2	20.1	64.0
2 SD	1.0	1.6	1.1	1.6	1.2	2.4	3.4	2.1	1.6	1.4

Note. All measures have been adjusted for scale and magnification to obtain life size values, and are in millimeters.

Discussion

Since frontal averages made from independent samples of 44 faces were perceived to be the same at rates well above chance and similar to control images (Fig. 39), averages made from 44 or more individuals appear to be reliable, at least for specific samples, e.g., individuals of a particular age, in a particular course, at a particular university, in a particular country. The observation that average faces made from independent samples of 65 faces were perceived as being the same more frequently than control images suggest that faces made from 65 individuals are highly reliable for such specific samples. To double check, a 130 face average was also constructed and tested against the 65 and 44 face averages using the sample of assessors above, although this 130 face average was not made from an entirely independent samples (e.g., when compared to a 65 face average only half the faces in the 130 face average were independent). The comparisons showed no large differences. Since recognition of faces as the same individual did not decrease when average faces made from 44 and 65 individuals were compared to the 130 individual average, it appears that using more than 65 individuals in an average will not increase the accuracy or reliability of (frontal or profile) averages made from highly specific samples much.

Although average faces made from 26 individuals were, in some instances, recognized at rates not statistically different from control images these faces seem not to indicate very reliable averages since they were not statistically above chance rates for guessing "same" or "different". Also perception rates of the 26 face averages as "being the same" was much less than those for 44 faces, which were much closer to control rates overall (Fig. 39).

While faces made from 10 individuals were not perceived to be the same, Table 22 indicates that metric differences between the faces were small. This suggests that even if a face has similar dimensions it may or may not be recognized as belonging to the same person. This is not unexpected, since it is well known that recognition decisions can be based on features like shape, colour and texture information rather than dimensional and configural characteristics (Fraser and Parker, 1986; Haig, 1984; Haig, 1986). This may have significant ramifications for the success of forensic facial approximations (a method used to rebuild faces from skulls), which presently relies heavily upon accurate replication of configural characteristics interpreted from the skull rather than feature specific cues (Taylor, 2001a).

There appears to be two alternative explainations for the decrease in "same face" responses for profile view (see Fig. 39). It may be becasue recognition in profile view is harder than in other views. Support for this is that response times are slower in profile (Bruce *et al.*, 1987). People may be more unsure when looking at profile views, and more likely to assume the faces are different when this is the case. However, it nust be noted that in Bruce's (1987) study participants were probably looking for evidence of "sameness" to base their decisions rather than evidence of "differences" (Bruce *et al.*, 1987), which may involve different processes (Bruce *et al.*, 1987). Since the average faces used in this study looked very similar, "differences" may have been used by participants in their decision-making rather than "sameness". If this was the case it may be that profiles views are more distinguishing than frontal views, rather than profiles being harder to recognize as being the "same". This issue may be resolved by conducting a study that determines if recognition response times for the same face sets (that include highly similar and different faces) differs depending on whether the question "are these faces different?" is asked in comparison to "are these faces the same?".

The findings of this research paper suggest that previous studies that have not used 44 or more individuals in their averaging sample may not provide reliable results since the average faces used are unlikely to represent the group from which the sample was drawn (an even less likely to represent broader population groups). This makes it difficult to repeat such studies because using a new set of individuals, even drawn from the same group, will not give repeatable face averages. Since this studies sample consisted of females between the ages of 18 and 34 years, who were predominantly "white" and studying dentistry, it is logical to conclude that the average faces generated here are not even representative all female dental students from the University of Melbourne between the ages of 18 and 34 years since not all dental students at The University of Melbourne are "white Europeans". This highlights further, the specific nature of averages that have been made to date, which are not general human averages, and should therefore be referred to specifically in terms of their sampling rather than just being called "human averages". Figure 40 shows examples of several average faces drawn from similar but more broadly defined groups (white European males), from different localities and averaged using calculated average methods. Despite a "common theme" between the images, it can be seen that the faces differ considerably. It seems likely that these faces would not be perceived as belonging to the same individuals and much of the similarity between the faces may be due to averaging techniques, e.g., the fuzziness of the hair and the smoothness of the image rather than actual face morphology.

(a) (b) (c) (d)

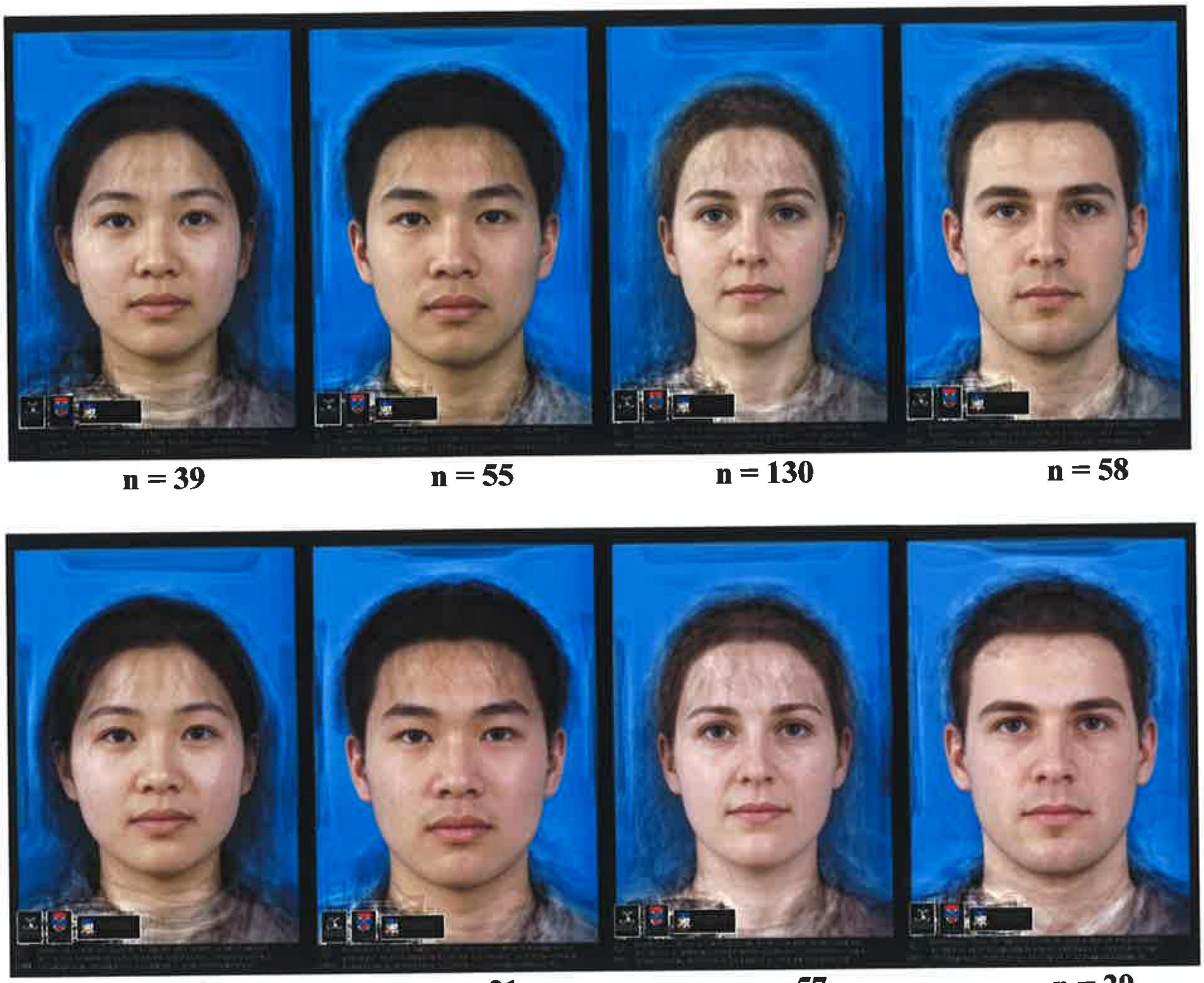
Figure 40: Comparisons of average "Caucasoid" males made from different samples, by different authors using similar techniques
Images have been standardized on interpupillary distance. **(a)** average (linear reverse) made by the authors using Psychomorph® and highly standardized photographs of 58 individuals, predominately dental students, aged 18-34 years, from The University of Melbourne, Australia. **(b)** average (linear reverse) made by Penton-Voak *et al.* (1999) using Psychomorph® and loosely standardized photographs of 21 second year undergraduates from The University of St. Andrews Scotland, mean age of 20 years. **(c)** male average made by Langlois using Gryphon's Morph® and "loosely" standardized photographs of 32 individuals from Texas, USA (courtesy of Langlois J). **(d)** male average made by using Gryphon's Morph®and "loosely" standardized photographs of 24 individuals from New Zealand and the US (reprinted with permission from Rhodes *et al.* (1999) © Blackwell publishers).

The specificity of human averages that have been made to date, may also affect the generality of research results interpreted from average face experiments, like those of preferences for facial attractiveness (e.g., Little *et al.*, 2001; Penton-Voak *et al.*, 1999; Perrett *et al.*, 1998; Perrett *et al.*, 1994; Rhodes *et al.*, 2000). For example, in studies of facial sexual dimorphism it is often assumed that the average female face is 0% feminine and the male face is 0% masculine, however since these averages are only sample means, the sexual dimorphism of these faces probably differs in relation to the sexual dimorphism of true group averages, that is, the average faces generated from the samples may in fact be more masculine or feminine than on average. Therefore it can only be concluded from face preferences studies that femininity / masculinity is attractive, not that masculine or feminine faces are attractive.

Since highly standardized photography techniques were used in this study the sample size needed to generate a reliable average (~45) is probably much less than in those instances where less standardized photography techniques are used because various photography conditions are unlikely to be random and average to the same mean, particularly in smaller samples e.g. n<50. Likewise, if sampling of individuals is not random, or includes individuals with large anatomical variability, the number of individuals needed to generate a reliable average is likely to be greater.

Since the sample sizes used to generate average faces for specific stereotypical population groups in the previous study were rather small we increased sample sizes by adding an additional 137 people to the averages overall. These additional faces were also taken on the same photography rig but by different investigators at different times. These average faces are presented in Figures 41 and 42.

It can be seen from Figures 41 and 42 that the revised average faces differ to some degree in shape, colour and texture to the original faces. Especially prominent are the colour differences, which may be due in part to slightly different photography conditions between the samples and possibly due to some degradation of slide films of earlier photographs during storage. Despite this, the revised averages are considered to be superior because they have been produced from larger samples. It is probably unrealistic to expect these faces to represent the average of all Australians, but since these are the only averages that have been developed for Australians using standardized photographs so far, they will be used in the next section for facial approximation on individuals from South Australia.



n = 39 n = 55 n = 130 n = 58

n = 28 n = 31 n = 57 n = 29

Figure 41: Revised average human faces (frontal)
Top Row: revised average faces made from larger samples. Bottom Rom: Original average faces made from smaller samples.

n = 39 n = 55 n = 130 n = 58

n = 28 n = 31 n = 57 n = 29

Figure 42: Revised average human faces (profile)
Top Row: revised average faces made from larger samples. Bottom Rom: Original average faces made from smaller samples.

Facial Approximation Recognition: Assessment And Potential

How to Assess Facial Approximation Accuracy: Resemblance Ratings versus Face Pools

This chapter formed the basis for a full-length paper publication in the *Journal of Forensic Sciences,* 2002, (Appendix 9).

Introduction

Since the ultimate goal of facial approximation is to promote specific and purposeful facial recognition (Prag and Neave, 1997; Taylor, 2001a), an accurate facial approximation should be easily recognized as the person to whom the skull belonged (target individual).

As previously mentioned, the accuracy of facial approximations has been assessed experimentally by face pool comparisons (Snow *et al.*, 1970; Stephan and Henneberg, 2001; van Rensburg, 1993). This method requires a facial approximation to be shown to a group of assessors who attempt to identify the target individual out of a number of other faces (the face pool). The face pool is usually made up of the target individual's face and non-target faces, of the same age, sex, and population of origin as the target individual. However, depending on the experimental method the target face may or may not be present in all face pools. Once the assessors have attempted to identify the target individual, the confidence at which identification rates for each face can be considered to be above, below, or equal to chance, can be calculated using statistical methods (e.g., chi squared test, Fisher's exact test). The higher the identification rate of the target individual above chance, at statistically significant levels, the more accurate the facial approximation.

The accuracy of facial approximations as measured by face pool comparisons appears to be generally low. Stephan and Henneberg (2001) found only one of 16 facial approximations to be identified at a statistically significant rate (25%) above chance ($p<0.05$), with many non-target individuals identified for all facial approximations, some significantly above chance rates ($p<0.05$). Snow *et al.* (1970) found two of two facial approximations to be identified significantly above chance ($p<0.05$), one identified 12% above chance rates, the other 54% above chance rates. Again, non-target individuals were identified in both cases. In 'case 3', two non-target individuals (photos 4 and 6) were selected at rates close to that of the target individual (Snow *et al.*, 1970) and appear to be above chance at statistically significant levels ($p<0.07$). Van Rensburg (1993) found that 15 facial approximations were, on average, identified at a rate 19% above chance, with the remaining identifications being of non-target individuals.

Not all authors have, however, assessed a facial approximation's accuracy by testing its ability to be recognized. Others (Diedrich, 1926; Helmer *et al.*, 1993; Krogman, 1946; Prag and Neave, 1997; Stadmuller, 1922; Stadmuller, 1925; Suzuki, 1973; von Eggeling, 1913) have attempted to assess the accuracy of forensic facial approximations by directly comparing the appearance of the facial approximation to the corresponding target individual for similarities (expressed as a resemblance rating).

The accuracy of facial approximations, as judged by recent direct comparisons to a target individual, appears quite high in comparison to the accuracy obtained in experiments previously mentioned. Examples of the judged accuracy by direct comparison are: [the facial approximation is] 'recognizable as that of the subject chosen' (Krogman, 1946); 'the resemblance between the two [the target individual and the facial approximation] was quite striking' (Suzuki, 1973); and 'The reconstructed face bore an uncanny resemblance to the photograph' [of the target individual] (Prag and Neave, 1997).

Helmer *et al.* (1993) measured the resemblance of facial approximations to their corresponding target individuals on a scale from 1 to 5 (1 being a great resemblance and 5 no resemblance). They concluded that 'in general it can be said that at least a slight (rating of 4) and often even a close resemblance (rating of 2) was achieved' (Helmer *et al.*, 1993). However, direct comparison assessment from older studies (Diedrich, 1926; Stadmuller, 1922; Stadmuller, 1925; von Eggeling, 1913) have found little resemblance between facial approximations and target individuals.

It appears that resemblance ratings are used to indicate the accuracy of a facial approximation because it seems that when two faces are similar they are recognizable. However, the validity of using resemblance ratings to assess a facial approximation's accuracy can be questioned for three reasons:

(1) Similar faces may not be the only recognizable faces. A face that does not appear to be morphologically similar to another may still be recognizable as belonging to the same person if observers are able to perceive recognizable characters in both faces despite their morphological differences. This may be evidenced in forensic casework where poor quality facial approximations, despite bearing limited resemblance to target individuals, are still identified correctly. Caricatures, as well as coarsely pixilated images of faces, present a similar scenario since these images are not precise representations of an actual face yet often remain recognizable (Benson and Perrett, 1991; Rhodes *et al.*, 1987). Furthermore, caricatures have actually been shown to increase the ease of recognition of familiar faces (Benson and Perrett, 1991; Rhodes *et al.*, 1987).

(2) Resemblance ratings are not a relative measure, that is, they do not take into account non-target faces, which may bear equal or higher resemblance to the facial approximation making them more recognizable than the actual target face.

These limitations may be the etiology of the discrepancy between the accuracy of facial approximation as measured by face pool comparison (low accuracy) and recent direct comparison methods (high accuracy). This study tests the second limitation by determining if facial approximations correctly identified as the target individual receive a higher resemblance rating than those facial approximations that are incorrectly identified as the target individual.

Methods

Four skulls were approximated with four different techniques of facial approximation: (1) a 3D American sculpting method; (2) a 3D combination sculpting method; (3) a 2D FACE assisted computer method; and (4) a 2D American drawing method (for details of techniques and examples of facial approximations see Stephan and Henneberg (2001)).

Thirty-seven assessors, with a background in the medical sciences, attempted to identify target individuals from a face pool for each facial approximation. Face pools consisted of 10 photographs. Antemortem photographs were used of the target individuals. Non-target faces in the face pools were of the same sex and approximate age as the target individual. Faces in the face pools were standardized for size, although this resulted in some photographs differing in resolution.

Since antemortem photographs of target individuals were used, the choice of photographs was limited and resulted in one photograph of a target individual wearing a hat and another sunglasses. In these cases the corresponding faces in the face pool also had similar attire e.g. hat or sunglasses. All photographs were developed and printed on Ilford®IS3.1M photographic paper (127mm x 100mm) in black and white. In order to keep lighting between the faces in the face pools and the facial approximations consistent, facial approximations were photographed in a fluorescent-lit room without a flash. This was done to simulate an average, indoor, amateur ‘snap shot’, which many of the photographs of the target individual faces appeared to be.

Assessors were presented with a facial approximation and a corresponding face pool (which included 10 faces) and asked if they could identify a face from the face pool that was the individual approximated. Assessors had the option of not being able to make an identification, that is, deciding that the facial approximation did not correspond to any face in the face pool. Not all face pools included the target individual. Of 592 identification scenarios, 472 included the target individual and 120 did not. Face pools that did not include the target face had one of the other faces in the face pool repeated in a slightly altered position so that no new individuals were introduced into the face array, keeping face pools as consistent as possible. Repeated faces were developed in black and white on a slightly higher contrasting (Ilford® IS4.1M) photographic paper (127mm x 100mm).

Facial approximations, with corresponding face pools, were presented to assessors who followed written instructions and completed a questionnaire regarding: which face (if any) the assessor could identify as being that of the person approximated; the resemblance of the facial approximation to the face identified by the assessor; and the confidence with which the assessor thought they had made a correct identification. As assessors completed one assessment scenario (one facial approximation compared to the corresponding face pool) they were given another (in random order) until all 16 assessments were completed. Since four different methods of facial approximation were used on each of the four skulls, face pools were repeated for each facial approximation of the same individual. Since assessors were not aware of the total number of faces approximated and that each face pool included only one target individual, assessors were forced to treat each identification assessment as an independent case.

The written instructions indicated that when an identification was made, assessors were to rate the resemblance of the facial approximation to the face identified on a scale from 0 to 10. Weighting of the assessors' judgements was guided by the description that: 0=no approximation; 2=slight approximation; 4=some approximation; 6=close approximation; 8=great approximation; and 10=perfect approximation. Every other number between 0 and 10 represented the mean weighting of the numbers immediately higher and lower. Assessors were also asked to rate the level of confidence that they had correctly identified the target individual by indicating if they were: not confident; slightly confident; fairly confident; or confident.

The distribution and variance of the two samples (true positive identifications and false positive identifications) were analyzed to determine if the assumptions for an unpaired t-test were fulfilled before a two tailed, equal variance, unpaired t-test was used to determine if any statistically significant difference existed between the average resemblance ratings ($p<0.05$, power=90%). Pearson's product-moment correlation coefficient was also used to determine if any relationship existed between the resemblance and confidence levels of both samples. Microsoft® Excel 98 was used for all data analysis.

Results

Of the total identification scenarios (592) there were 151 instances (26%) where no identification was attempted and 441 instances (74%) where identifications were made. Of these identifications, 354 (80%) were made when the target face was present in the face pool and 87 (20%) when the target face was not present. Thirty-eight (9%) of the identifications made were true positive (correct identifications of the target individuals) and 403 (91%) were false positive (identifications of non-target individuals).

A histogram indicated that both the true positive and the false positive samples were normally distributed, i.e., 68% of sample scores fell within one standard deviation of the mean and 99% of samples scores fell within three standard deviations of the mean (Table 23). An F-test indicated that variance of the average resemblance ratings for correct and incorrect identifications of target individuals were not significantly different ($p<0.05$). A two tailed, equal variance, unpaired t-test ($p<0.05$) indicated that there was no significant difference between the average resemblance for correct and incorrect identifications of target individuals at a power of 90% (Fig. 43).

The resemblance rating of the facial approximation to the face identified tended to increase as the assessor's confidence level increased, independent of whether or not the correct face was identified (Fig. 44). These trends were significantly correlated ($r=0.95$). Large standard deviations for resemblance ratings (as reflected in Figs. 43 and 44) and large ranges in confidence intervals were also observed.

Table 23: Percentage frequency of resemblance ratings for true positive and false positive identifications

Resemblance	True Positive (%)	False Positive (%)
0	0.0	0.2
1	0.0	1.5
2	10.5	6.9
3	2.6	8.7
4	10.5	13.9
5	31.6	20.8
6	18.4	19.1
7	18.4	17.9
8	7.9	10.2
9	0.0	0.5
10	0.0	0.2
Average (SD)	5.3 (1.7)	5.3 (1.8)

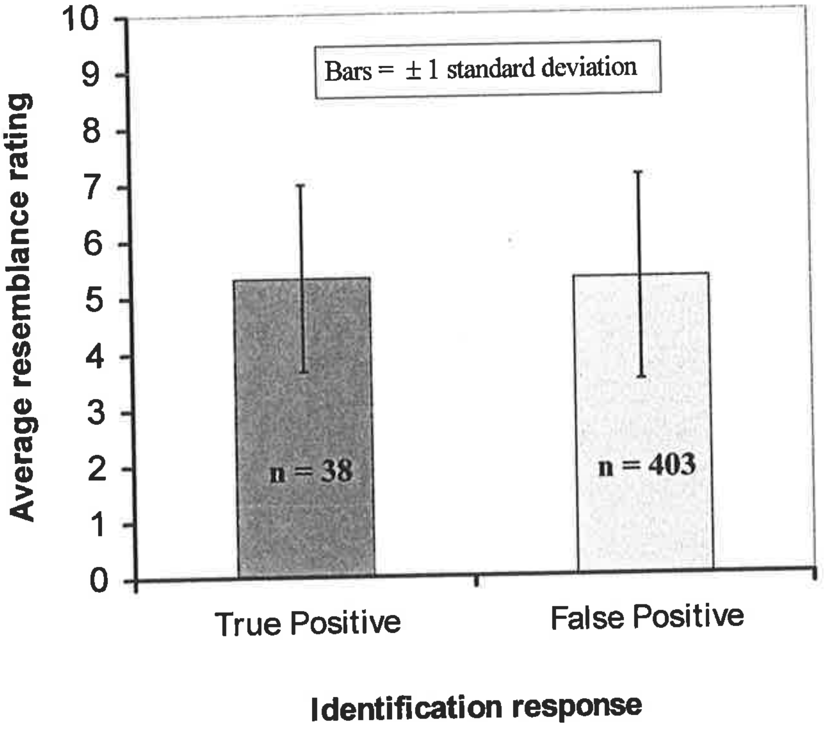


Figure 43: Average facial approximation resemblance for true positive and false identifications of faces.

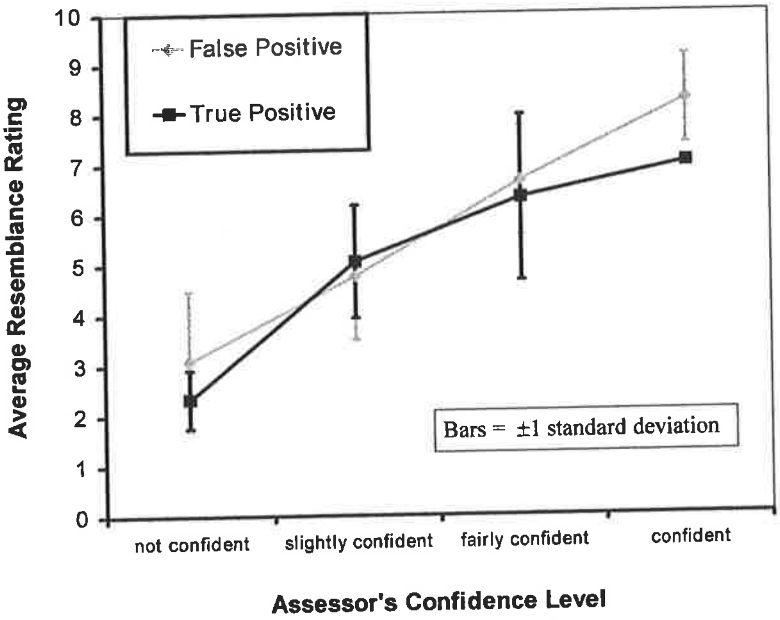


Figure 44: Average resemblance ratings by assessors' confidence levels comparing true positive and false positive identifications.

Discussion

The observation that average resemblance ratings for true positive and false positive identifications do not differ at statistically significant levels ($p<0.05$, $ß=0.10$) does not support the thesis that resemblance ratings indicate a facial approximation's accuracy (ability to be recognized as the target individual). It can, therefore, be concluded that resemblance ratings of a facial approximation to a target individual, *do not* indicate a facial approximation's accuracy.

The increasing resemblance rating with the assessor's confidence level is to be expected since it is logical to assume that when the resemblance of a chosen individual's face to a facial approximation is perceived as being great, the assessor would believe he/she has made the correct choice and be confident of it. However, the results indicate that an assessor may not have made a correct identification even if an assessor is highly confident s/he has selected correctly and believes that the resemblance is high. This finding is also consistent with the results of police lineup studies (Bothwell *et al.*, 1987; Cutler and Penrod, 1995; Lindsay *et al.*, 1998; Sporer *et al.*, 1995). The large range in responses for resemblance and confidence levels is not unexpected since both are subjectively determined.

Since resemblance ratings appear not to be valid in assessing the accuracy of a facial approximation, articles that have used such methods to assess the accuracy and/or quality of facial approximations (Gerasimov, 1971; Helmer *et al.*, 1993; Krogman, 1946; Prag and Neave, 1997; Stadmuller, 1922; Suzuki, 1973; Taylor, 2001a; von Eggeling, 1913) are probably unreliable. Future studies assessing the accuracy of facial approximation should use the face pool comparison as opposed to the direct comparison method (resemblance ratings).

This study also demonstrates that printing images of a facial approximation and the corresponding target individual (as has been done in *almost all previous* publications of facial approximation) to indicate to readers a facial approximation's accuracy is of little use. In such a scenario, only the similarity between the two is indicated, not the "recognizability"/accuracy of the facial approximation.

Since face pool comparisons assess a facial approximation's ability to be recognized, those studies that have used this method appear to give the best indication of a facial approximation's accuracy and quality. The main disadvantage of this method is that only an unfamiliar identification scenario (the use of assessors who are not familiar with the target individual) has previously been employed. Unfamiliar identification scenarios are not representative of a real forensic environment because a familiar person usually recognizes the facial approximation. It has been shown that recognition of unfamiliar faces is much poorer than is the case with familiar faces (Hancock *et al.*, 2000). Unfamiliar face identification from still video images in a face pool comparison scenario of 10 faces shows that approximately 30% of responses can be expected to be incorrect, even if viewpoints of the faces and expression are constant (Hancock *et al.*, 2000). Therefore, assessing a facial approximation's accuracy in a familiar identification scenario may increase the accuracy estimates of facial approximation.

Unfamiliar identification scenarios are also not representative of a real identification scenario since in familiar recognition it is the internal features of the face, as opposed to the external features, that play a more significant role in recognition (Ellis *et al.*, 1979). For unfamiliar scenarios the external features of the face appear to contribute as much as, or more than, the internal features for recognition (Ellis *et al.*, 1979; Hancock *et al.*, 2000). Present understanding of face recognition also shows differences in the mechanisms used to recognize a face. In an unfamiliar identification scenario recognition is primarily based on pictorial codes (Bruce and Young, 1986). In comparison, recognition of faces in familiar identification scenarios appears to be based on structural codes (Bruce and Young, 1986). The use of structural codes in unfamiliar scenarios may be facilitated however, by presenting assessors with several views of each photograph or by presenting all faces in the same exact conditions (e.g. same orientation, lighting, emotion etc.) (Bruce and Young, 1986).

Since achieving a familiar identification scenario for deceased individuals is difficult (Stephan and Henneberg, 2001), the face pool comparison method appears to be favorable as it presents assessors with a recognition task (despite the limitations listed above). However, the face pool comparison test is expected to be a more rigorous test of facial approximation accuracy than would probably occur in real scenarios since the use of structural codes is limited in face pool comparison methods. It is worth noting that if viewpoint or expression changes between the faces in a face pool or the stimulus face, misidentifications are expected to increase (Hancock *et al.*, 2000). Also, short video clips in comparison to still images, and colour images in comparison to gray scale, appear to offer no advantage, as they seem not to alter identification rates (Hancock *et al.*, 2000).

Ceiling Limits of Recognition From Two-Dimensional Facial Approximations

Introduction

Facial approximation methods rely heavily on averages for face construction and while this approach is useful (although not necessarily optimal (Simpson and Henneberg, 2002)) to maintain objectivity in methods, it is unknown if this method can actually result in specific and purposeful facial recognition. Therefore, any confidence in the "average approach" to facial approximation seems premature, as facial approximations made using these techniques may not be easily recognized.

The ability for facial approximation to achieve specific and purposeful facial recognition from average guidelines may be unlikely because facial approximation aims to predict unique individual appearances not average ones (Brues, 1958; Simpson and Henneberg, 2002). Also, when using averages to estimate a feature, prediction error is increased for many individuals if trait/s are not normally distributed, e.g., display binomial distributions. Average methods also increase the similarity between facial approximations and, therefore, probably makes specific recognition of facial approximations more difficult. Furthermore, the application of average methods to less typical, or more distinct, skulls decreases the distinctiveness of the face and makes those faces harder to recognize (Bruce and Young, 1998).

It is possible to determine the ceiling recognition rates for average facial approximations methods by representing features that are unlikely to ever be determined from the skull, e.g., specific face colour and texture, as averages while other facial characters, that may be potentially determinable from the skull, e.g., major face feature dimensions and shapes, are exactly representated. By doing this, this study aims to test the "best case scenario" or ceiling level recognition rates for average two-dimensional facial approximation methods by testing if facial approximations with average surface colour and texture but with exact face shapes (outlines) can be recognized. This test is useful because it will demonstrate the best case scenario for specific and purposeful recognitions from average facial approximation methods. It is also useful because the recognition rates found here will allow for the expected recognition rates of traditional (subjective) facial approximation methods, which are not exact, to be relatively gauged.

Methods

Average face generation

The one hundred and thirty female average face generated from people self-reporting a socially perceived group membership as "European" (mean age 22.3, sd 4.0 years), were used in this study. The average face template and the average face were generated as described earlier, p. 126 to 145 and p. 158-160. For convenience images are repeated in Figure 45.

(a) (b)

Figure 45: Average socially percieved European female face
(a) the calculated average face shape to which individual face shapes are warped and colour information blended (this delineation map is the same as that in Fig. 34).
(b) the final average face after warped faces have been blended together (this image is the same as the European female average face generated in Fig. 41).

Warping of the average face

Three other female individuals, self-identifying Europeans, from Melbourne Australia, were photographed and delineated as mentioned above. The average face was then warped to the exact same face shape for each of these three individuals using the Psychomorph® software and delineations of the faces (Fig. 46b). Figure 46a and 46c show the original and the warped average faces (mock facial approximations displaying exact face shapes but average colour and texture information) for each target individual respectively. Two other mock facial approximations were generated according to methods described above from randomly selected face photographs to test for false positives. These individuals were also from Melbourne, Australia. Thus, five stimulus mock facial approximations were generated in total including three whose target individuals were used in this study (to test for true positive recognitions) and two whose target individual faces were not used (to test for false positive recognitions). Two recongnition trials were conducted to test for recognition. One was an unfamiliar scenario, e.g., assessors did not know any of the individuals used in the study; the other was a familiar scenario, e.g., assessors knew the target individuals used in the study.

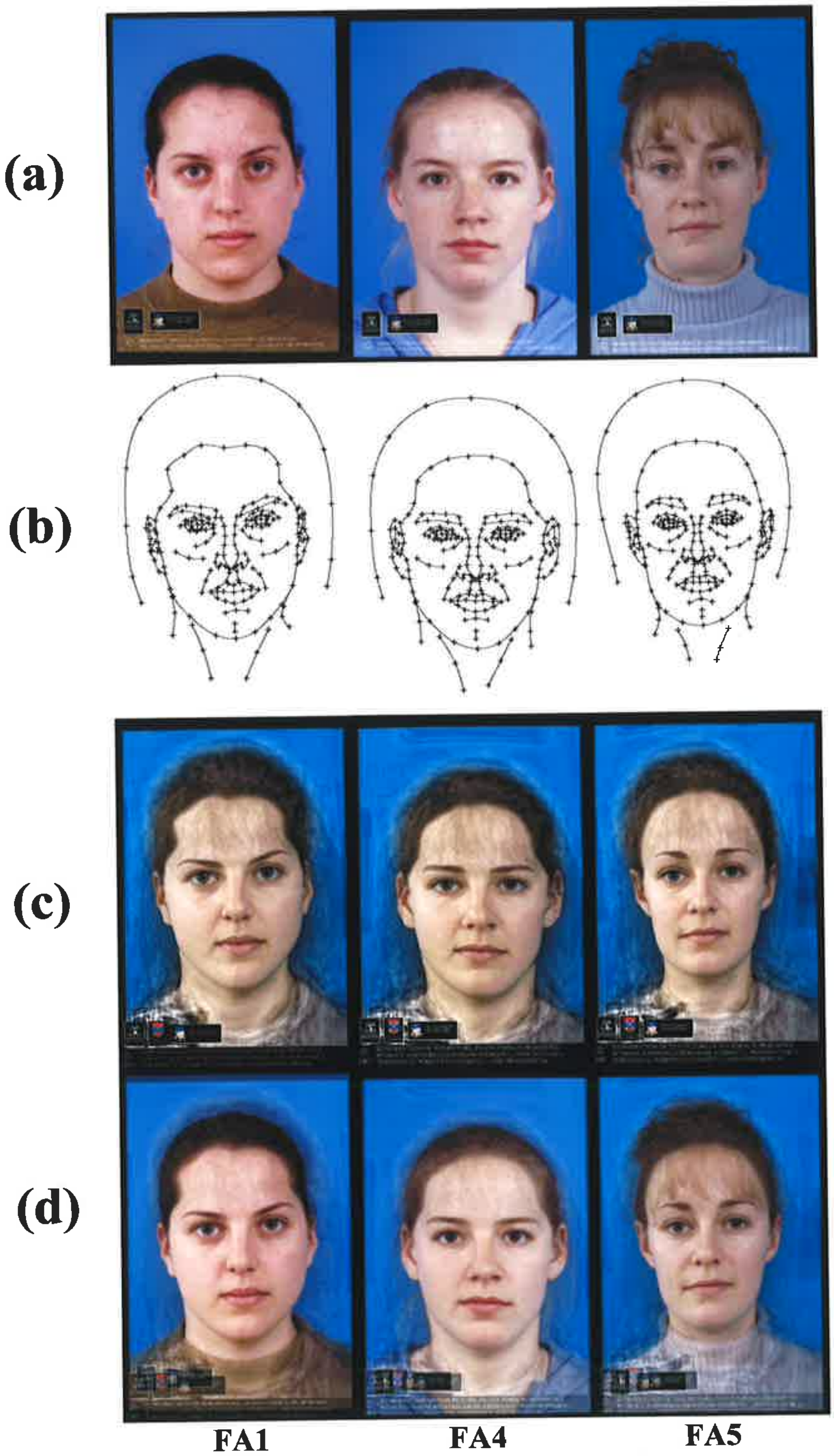


Figure 46: Generation of the mock facial approximations whose target individuals were used for true positive recognition tests
(a) illustrates the actual photographs of target individuals. **(b)** shows the delineations of face shapes for each individual. **(c)** shows the final mock facial approximations made by warping the colour and texture information of the average face (Fig. 45b) to each individual's face shape. **(d)** shows the superimposition of the facial approximation with the original target face to demonstrate that exact face shapes were replicated.

Unfamiliar recognition trials

Forty-three assessors (27 females, mean age 22 years, sd 8 years; 16 males, mean age 30 years, sd 14 years) from Adelaide, Australia, who were unfamiliar with all faces, i.e., did not know any of the people whose faces were used in the experiment, took part in five recognition trials (one for each facial approximation). Assessors attempted to identify from a face pool the target individiual whom the facial approximation represented. The face pool included 6 individuals from Melbourne, Australia (three target individuals and three non-target individuals, Fig. 47). The face pool was printed in colour on Epson® Photo Quality Ink Jet Paper (matt) using an Epson® Stylus® 890 printer. Each face image measured 100 x 66.7mm.

Figure 47: The face pool used in the unfamiliar identification tests.

Since the experiment was primarily run during a University open day, trials were conducted, one at a time, in a random set order (Facial approximation 1 first, then 2 then 3 etc., as indicated in Fig. 46), removing the need to randomly shuffle pictures and increasing the speed of participant throughput. We therefore traded time for possible trial order effects, although such effects may be reduced since only 5 trials were conducted and no participant appeared to express fatigue. The three facial approximations with "targets" were numbers 1, 4, and 5; while the two "non-target" facial approximations presented to test for false positives were numbers 2 and 3. Assessors were asked to examine a warped average face, compare it to the face pool, and attempt to make an identification, or indicate if they thought the target face was not present in the face pool. Assessors therefore had a 50% chance indicating that the target face was/was not present and an 8.3% chance of choosing any face in the face pool (50% chance that the face was there and a 16.5% (1/6) chance of choosing the right face from the six displayed; 50% x 16.5% = 8.3%).

A common face pool was used across recognition trials, so that the experimental procedure was simple, easily achieved by participants without constant investigator supervision, and so assessor throughput was rapid. This protocol, as opposed to using new face pools for each trial, seemed valid since assessors were aware that facial approximation methods included prediction error and that it was possible that two facial approximations while looking different may actually be of the same person. Also, since participants were unaware of how many different facial approximations (average facial warps) there were and if their respective target individual was, or was not, present in the face pool, they were forced to treat each trial independently. Despite the use of the same face pool several individuals choose the same faces as those chosen in previous trials confirming the validity of this approach (see results). Data were collated in the JMP 4.0 statistical package and analyzed using Fisher's exact and Chi-squared tests ($p<0.05$).

Familiar recognition trials

Ten assessors (mean age 22 years, sd 3 years) who knew at least one of the *target* individuals personally also completed the recognition trials, however, these differed from the unfamiliar trials since face pools were not used. Assessors were presented with each mock facial approximation (FA1, FA4, FA5) and asked if they recognized whom the facial approximation represented. A correct identification required participants to correctly name the person whom the facial approximation represented. Sample sizes of familiar assessors were: facial approximation 1: n=4; facial approximation 4: n=5; facial approximation 5: n=7.

Nine assessors (mean age 30 years, sd 10 years) who knew at least one of the *non-target* individuals included in the face pool used in the unfamiliar trials (face pool photo 2: n=2; face pool photo 4: n=7; face pool photo 6: n=1) also completed the familiar recognition trials to help test for false positives. After the recognition trials were completed, all assessors indicated how long they had known the person in question that they were familiar with and what their relationship with that person was (e.g., brother, friend etc.).

Results

Unfamiliar recognition

A considerable proportion (40%) of assessors identified the same face from the face pool, on one or more occasions, for different facial approximations (four assessors repeated identifications for two face pool photos), indicating that use of the same face pool did not result in assessors discounting a previously selected face for later trials. This suggests that original chance recognition rates appeared to be valid for later trials. Overall, there were 21 instances of repeated face selection in a total of 215 recognition trials. Face pool photo 1 (FPP1) was identified twice by 12 independent assessors; FPP2 by 1 assessor; FPP3 by 3 assessors; FPP4 by 1 assessor; FPP5 by 1 assessor; and FPP6 by 3 assessors. Facial approximations 3 and 5 (Fig. 48) were the most frequently identified (n=8) as representing the same face pool photo (photo no.1).

FA5 **FA3** Target Individual of FA5 (Face Pool Photo1)

Figure 48: The facial approximations frequently identified as the same target individual in unfamiliar tests.

Of the 129 identifications made when the target face was present in the face pool, 55 (43%) were true positive (correct) identifications, 36 (28%) were false positive (incorrect) identifications and the remainder consisted of 38 (29%) instances where no identification was (incorrectly) made. Two of the three target faces were correctly identified above chance rates, $p<0.05$ (Fig. 49), the other being close to chance, $p<0.07$. Facial approximation no. 5 (FA5) was identified the most accurately, with identification rates 56% above chance rates (Fisher's exact test, $p<0.001$). FA1 was identified 32% above chance rates (Fisher's exact test, $p<0.001$). FA4 was identified the least accurately at 16% above chance rates (Fisher's exact test, $p<0.07$). These results indicate that facial approximations displaying average colour and texture may be correctly identified by unfamiliar assessors, but rates were not very high (average of ~43% above chance), even though most were statistically significant.

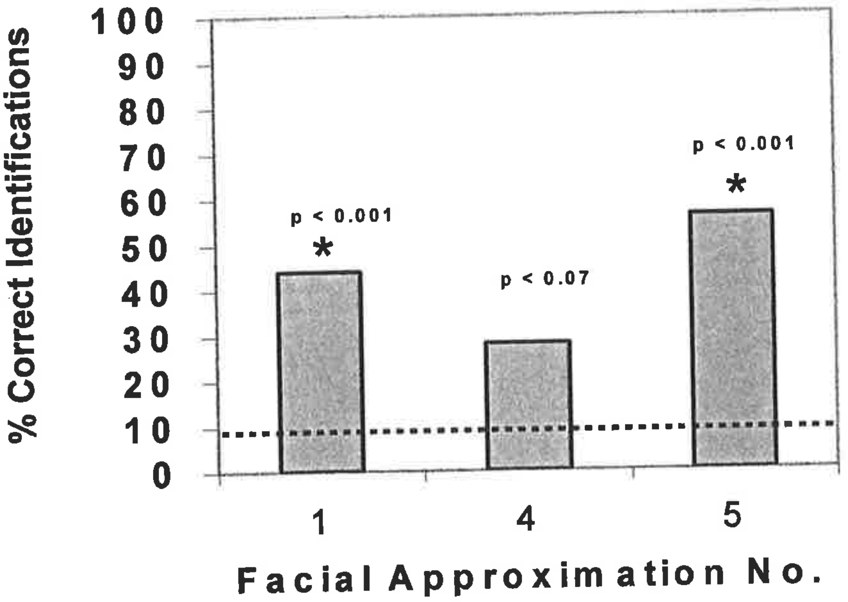


Figure 49: True-positive identification rates of facial approximations with target individuals

------- indicates chance rate; * indicates statistical significance

In the two instances where target faces were not present (FA2 & FA3) identifications of the target face as not being present in the face pool were not significantly different from chance rates (Chi-squared test, $p<0.05$, Fig. 50). This indicates that even though participants could correctly identify target individuals when present they could not tell when target individuals were not present.

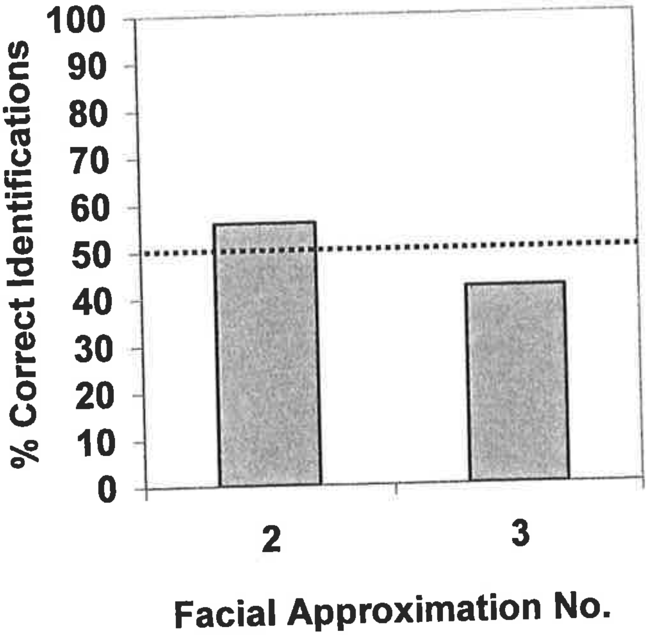


Figure 50: False-positive identification rates of facial approximations without target individuals
------ indicates chance rate

Familiar recognition

Of the three facial approximations represented by target individuals in the face pool (i.e., FA1, FA4, FA5) two were correctly identified. Of the seven people familiar with the person represented by FA5, six made correct identifications (86%). Most of these assessors were colleagues of the target who had known her for three years. Of the five people familiar with the person represented by FA1, two made correct identifications (40%). Most of these assessors were friends of the target and had known her for about four years. It is worth noting that a sibling of the target of FA1 who had known her for almost 20 years did not make a correct identification. Of the four people familiar with the person represented by FA4, none made correct identifications (0%). Most of these assessors had known the target for about two years but again a sibling who had known the target for almost 20 years also did not make a correct identification.

Only two false positive identifications were made out of a total of 95 familiar identification responses (2%). One assessor in the familiar testing scenario correctly identified one of the "non-target" facial approximations, which was not represented by its respective target face in the face pool used for unfamiliar testing.

Comparison between testing scenarios

Comparison of identification rates in unfamiliar trials to familiar trials indicated that both follow similar trends, however, the magnitude of responses differed. For example, in both familiar and unfamiliar scenarios FA5 was correctly recognized the most, followed by FA1 and then FA4. However, FA5 was recognized considerably more reliably in the familiar identification (86%, n=7) compared to the unfamiliar identification scenario (56%, n=43). FA1 was recognized at similar rates in both scenarios, 40% (n=5) in familiar compared to 44% (n=43) in unfamiliar scenarios. FA4 was recognized much less in familiar scenarios (0%, n=4) compared to unfamiliar scenarios (28%, n=43). Overall, average recognition rates were similar between testing scenarios being 42% for familiar, and 43% for unfamiliar. There were considerably less false positive identifications in the familiar testing scenarios (2%) in comparison to the unfamiliar tests (28%).

Discussion

The results from both unfamiliar and familiar recognition trials generally indicate that two-dimensional average faces warped to individual face shapes can be used to promote facial recognition of the majority of target individuals, however, recognition rates are generally low (average of 43%). This suggests that the average approach to facial approximations may work so long as an individual's face shape can be very accurately predicted. Further tests are needed to determine at what stage deviations from individual face shapes cause significant reductions in recognition accuracy.

This study demonstrates a best-case scenario for recognition of facial approximations made using average methods since exact face shapes were replicated here. This is not possible in forensic casework. The recognition rates observed in this experiment are, therefore, expected to be *dramatically* higher than those that would be obtained in general facial approximation practice. This appears to be the case since those who have conducted scientific tests of traditional methods (Snow *et al.*, 1970; Stephan and Henneberg, 2001; van Rensburg, 1993) generally report rates that are less than 48% above chance (the highest rate obtained in this study using unfamiliar tests). Although the reliability of facial approximation in this study was rather high, i.e., facial approximations recognized above chance in four out of six trials (five out of six trials if the borderline statistical significance is included), the reliability in practice for traditional methods is probably much less since exact face shapes cannot be represented like they were here. This study, therefore, supports conclusions of others that traditional facial approximation methods are probably inaccurate and unreliable (Stephan and Henneberg, 2001).

Even though true positive recognitions were generated in this experiment, recognition rates were low, for example, the highest generated in the unfamiliar scenario was 48% above chance which indicates that if one person said they recognized the facial approximation no reliable conclusion could be made if they were right or wrong. However, all facial approximations were identified at least once, except FA4 in familiar tests, indicating that most had the potential to be identified in a forensic environment. It should be noted, however, that successful recognition in the laboratory environment does not necessarily mean that the facial approximation will be recognized in the field since laboratory tests exclude additional variables like the broadness of media coverage/who sees the media reports etc. (Haglund, 1998).

The finding that trends seen in the unfamiliar trials were similar to those in familiar scenarios suggests that unfamiliar testing scenarios are useful in evaluating facial approximation accuracy even though facial approximations are usually recognized by familiar individuals, not unfamiliar ones. This is significant for laboratory tests of facial approximation since testing of familiar individuals can be problematic due to availability of such individuals and ethical concerns for their prolonged exposure to traumatic events. One limitation to unfamiliar testing, indicated by this study, is that false positive identifications appear to be significantly increased in comparison to familiar tests. Therefore, suggestions that small but significant recognition rates of facial approximations in unfamiliar scenarios (e.g., 8%) may be troublesome since many significant false positive identifications may be generated (Stephan and Henneberg, 2001), does not appear to be true for forensic environments where familiar individuals usually recognize the facial approximation.

The problem of false positive recognitions in unfamiliar tests may be resolved by using sequential lineups rather than face pools since this procedure has been shown to decrease false positive identifications without significantly effecting true positive identification rates for eyewitness identification (Lindsay *et al.*, 1991a; Lindsay *et al.*, 1991b; Lindsay and Wells, 1985; Steblay *et al.*, 2001). In sequential lineups (face pool) images are shown one at a time to the assessor who decides wether or not that person is the target individual before the next person in the sequence of face pool photographs is shown (note that the experimenter holds more photographs than is to be shown so that the assessor cannot anticipate the end of the face array). The reasoning behind this is that assessors are forced to rely more on absolute criteria (e.g., is this the target individual or not?) rather than realtive criteria (is this individual more like the perpetrator than other lineup members?) (Lindsay and Wells, 1985). Such an approach may be useful to the assessment of facial approximation also, however, it is worth noting that facial approximations may be so unrepresentative of target individuals that assessors find it difficult to make absolute responses and consequently no identification responses may result. This would be a good additional test to further establish how accurate facial approximations actually are.

It may be possible that recognitions could be established from face outlines (e.g., face templates) alone without the need for average colour and texture information to be addtitionally presented. However, recognition rates of studies testing face outlines appear to be less than those obtained in this study (average 43%) when average colour information was also included indicating that the inclusion of this information is useful. Rhodes et al. (1987) found that frontal face outlines alone, similar to those used here to generate the averages but using 169 points rather than 219 points, were recognized 38% of the time. Davies et al. (1978) found that their "line outlines" (which displayed slightly less information than the face delineations used in this study) were recognized at a rate of 23%, while their line drawings (which displayed much more information than the face delineations used here) were recognized 47% of the time. Therefore, there seems to be some advantage in using average colour in 2D facial approximations since recognition rates appear to be increased slightly (~5%), perhaps because the faces appear more realistic. This may not be surprising since it is known that slow-varying intensity patterns (colour and texture information) contribute to facial recognition (Bruce and Langton, 1994; O'Toole *et al.*, 1997) and the inclusion of average colour and texture would crudely approximate that of many individuals. While it is unknown if face delineations alone can generate recognition rates equal to that of warped average faces, even if they did it would not argue against the use of average colour and texture since this objectively represents faces unlike traditional two-dimensional facial approximation methods where face colour and texture are subjectively represented by pencil toning.

As previously mentioned, average colour and texture may also be useful for three-dimensional computer approximations where faces are derived by warping laser scanned images of heads to average soft tissue depths placed on skulls (Vanezis *et al.*, 1989; Vanezis *et al.*, 2000). Since the recognition of such laser scanned faces, is well below that of photographs (Bruce *et al.*, 1991) it is possible that their recognition may also be increased (perhaps only slightly) by the inclusion of average colour and texture information. Additionally, average faces seem advantageous to practitioners for three-dimensional clay facial approximation because they can be used as a basis for constructing the face maintaining objectivity to a greater degree.

An interesting observation of this study is the similarities/differences between the two facial approximations (FA3 and FA5) perceived by many unfamiliar assessors to represent the same individual (FPP1). Visually, the face shape of these two facial approximations is rather different (Fig. 48), yet they were recognized as representing the same person. FA3 appears much more euryprosopic and brachycephalic than FA5 (which exhibits the exact face shape of the target individual). Additionally, the shape/position of the upper lip, eyebrows and ears differ considerably. However, both facial approximations display similar hairlines that deviate to the right, left eyes that are positioned slightly higher than the right eye, and rather "square" jaw lines. Perhaps these were the features that resulted in the faces being identified as the same person by unfamiliar assessors. Either way, since no familiar assessor incorrectly identified FA3 as belonging to the target individual of FA5 it appears that familiar assessors were able to differentiate between these facial approximations, unlike unfamiliar assessors.

Applying Knowledge Described Here To Facial Approximation Methods: Does it Increase Their Accuracy?

Introduction

It has been found in the preceeding sections that several traditional methods for building prominent facial features from the skull have considerable inaccuracy, like those for eyeball projection, pronasale position, mouth width and superciliare position. New guidelines that better predict these soft tissue features have also been presented as well as methods that objectively represent features that cannot yet be determined from the face, like average face colour and shape. Additionally, it has been demonstated, by using exact face shapes and average face morphology that the ceiling limit of two-dimensional facial approximations is about 48%, and that face pools should be used for assessing the accuracy ("recognizability") of facial approximations. This aim of this study is, therefore, to determine if by using the average faces and the new improved soft tissue prediction guidelines described here facial approximations can be generated which are correctly recognized above chance rates as tested in an unfamiliar identification scenario.

Although the improvements made to traditional soft tissue prediction guidelines made above are significant, they are few and hence close replication of individual face shapes and increases in recognition rates may be unlikely at this stage. However, it would be encouraging if recognition rates were found to increase since this would indicate immediate rewards for the use of a few systematically tested guidelines. Even if recognition rates are not found to increase, those determined will be useful since they will enable traditional subjective methods of facial approximation to be relatively judged.

Methods

Facial approximations were constructed for each of three skulls ("Kate": female, aged 20-30 years, European population of origin; "Jane": female, aged 30-40 years, European population of origin; and "Fred": male, aged 40 years, European population of origin) previously used for facial approximation testing by Stephan and Henneberg (2001). The skull of "Sam" was not included since his age (40-50 years) fell well outside that of individuals used to generate the average faces (mean age 22 years, age range 18-34 years). "Fred" was also outside this range but not by as much as "Sam" and was therefore included to increase the sample size. "Jane" was borderline with respect to age (30-40 years) so she was included. "Kate", who was aged about 20-30 years was well suited to the average face age range. The use of these skulls was convenient because results of tests conducted here using new facial approximation methods could be compared to those of Stephan and Henneberg (2001) using traditional subjective methods. Since the new methods employed here are objective (because they use average faces and predetermined soft tissue prediction guidelines), the fact that I had previously seen the actual facial appearance of target individuals was not considered to be a significant factor since any bias in face construction, i.e., deviations from predetermined soft tissue prediction guidelines, would be evident in the final facial approximations.

Construction of the facial approximations

The facial approximation method used here is best classified under the American technique since it does not rely on any facial muscle buildup. This is probably an advantage since determining facial muscle origins, insertions and shape/bulk is highly subjective as these muscles leave little clue to their form on the skull as previously mentioned. The technique used here is an adaptation of the averaging/caricaturing software developed by Perrett and colleagues (Benson and Perrett, 1991; Rowland and Perrett, 1995; Tiddeman *et al.*, 2001) at the University of St. Andrews. The technique requires the delineation of the two-dimensional face shape over the skull using a large number of landmarks joined by contour lines. Landmarks determining the face shape are located using as many tested soft tissue prediction guidelines as possible with other unknown aspects being supplemented with untested guidelines. Totally unknown feature shapes are left as much as possible to the computer normalized shape of the average face which is based on three reference points (see below), hence limiting the subjectivity of determining unknown positions of certain landmarks. Some landmarks required small amounts of "tweaking", to ensure contour lines did not overlap or were too close since this can produce artifacts in final images (see e.g., Fig. 51). The sex, age, and population specific average colour and texture information was then warped to the delineated face shape to obtain the final facial approximation.

(a) (b) (c)

Figure 51: Averaging artifacts caused by contour lines too close or overlapping

(a) delineation template with overlapping neck/gonion region, inappropriately placed distal nasolabial fold points, and overlapping proximal epicanthal points. **(b)** resulting face warp with distortions; note gonioal region abberations, exaggerated lines at corners of mouth (oromental grooves), and abnormal medial eye morphology. **(c)** face warp created using a delineation template with non-overlapping contour lines.

Skull casts were used as the basis of facial approximation so small-diameter, sharpened plastic rods could be rigidly mounted at realistic angles. A heated sheet of modeling wax was used to ensure a free way space of about 2mm (Fig. 52), which naturally occurs between the occlusal surfaces of the teeth when individuals are relaxed (Nairn, 1976). Additionally, the interarticular disc of the temporomandibular joint was also simulated using modeling wax. Age specific soft tissue depths reported by Helmer (1984) using ultrasound were used to establish general face shapes. Only depths at landmarks thought to be useful with regard to the new methods were used (Table 24). Mid-sagittal depths were oriented at realistic angles according to guidelines of George (1993).

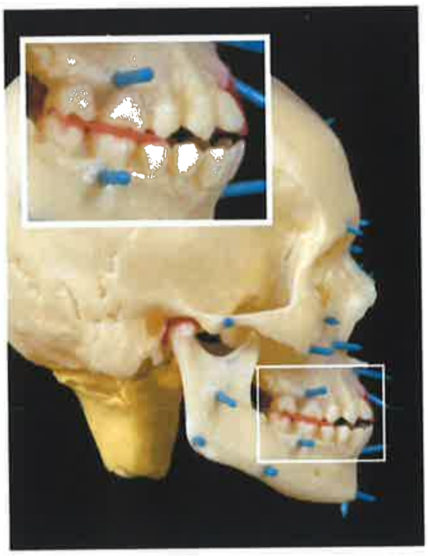


Figure 52: Use of heated wax to achieve free way space and hence the correct position of mandible in relaxed state
Also note the simulation of the interarticular disc of the temporomandibular joint using modeling wax.

Skull casts were mounted in a typical natural head position, and photographed at a glabella – film plane distance of approximately 1204mm with a 35mm SLR camera fitted with a 105mm lens, in both frontal and right-profile views. Photographs were then scanned into a computer using a flat bed CanoScan D 660U scanner and images were saved as JPEG's. All measurements made from the images for purposes of facial approximation were adjusted for scale according to values of direct measurements taken on the actual skull casts.

Table 24: Selected average soft tissue depths of Helmer (1984) used for facial approximation
All measures have been rounded to the nearest 0.5mm.

Location	Number	Use	Kate	Jane	Fred
			Female, 20-29yrs	Female, 30-39yrs	Male, 40-49yrs
Vertex	1	frontal/profile	4.5	5.0	5.0
Euryon	29	frontal	5.0	5.5	6.5
Zygion	31	frontal	5.0	5.0	5.5
Mid-masseter (between gonion and zygion)	32	frontal	17.0	18.5	20.5
Gonion	33	frontal	11.5	11.5	13.5
Gnathion	16	frontal/profile	7.0	7.0	9.5
Metopion	3	profile	4.5	4.5	5.0
Glabella	5	profile	5.5	5.5	6.0
Nasion	6	profile	7.0	6.5	7.0
Mid-nasal	7	profile	3.0	3.0	4.0
Rhinion	8	profile	2.5	2.5	2.5
Subnasale	11	profile	14.0	13.0	15.5
Labrale superius	12	profile	12.0	10.5	12.5
Labrale inferius	13	profile	12.0	12.0	14.0
Mentolabial sulcus	14	profile	10.5	11.0	13.5
Pogonion	15	profile	9.5	10.0	11.5

Frontal Facial Approximation of specific features:

General face outline

The 2D average face shape was initially adapted to the skull in frontal view using the pupil centers (landmarks 1 & 2) and stomion (landmark 1) as reference points (see Fig. 32 or Appendix 6 for all landmark positions named in this section). Adjustments were then made to the normalized 2D face shape delineations, as required, to accommodate the underlying skull with average soft tissue depths (Fig. 53 to 58). Relaxed average faces were used since skull casts were mounted with a freeway space representative of relaxed facial posture.

Eyes

The pupil centers (delineation landmarks 1&2) were positioned according to methods of Eisenfeld et al. (1975) (Fig. 53 to 58). This method was used above untested guidelines of Krogman (1962), and Gatliff and Taylor (Gatliff, 1984; Gatliff and Taylor, 2001; Taylor, 2001a) because they have at least had some empirical validation. Iris diameter was constructed as 12mm as suggested by Taylor (2001b). Medial (landmarks 20 & 28) and lateral canthi (landmarks 24 & 32) positions were determined according to methods of Krogman (cited in Caldwell, 1986).

Eyebrows

Superciliares (landmarks 111 & 119) were located using the very general guidelines reported by systematic tests in this thesis (p. 120 to 121). Eyebrow shape was left to computer normalization of the average face shape as much as possible.

Nose

Alars (landmarks 57 & 67) were positioned according to nasal aperture 3/5ths rule as tested by Hoffman and colleagues to determine nose width (Hoffman *et al.*, 1991). Subnasale (landmark 62) was positioned below the nasal spine according to either the method reported by Schultz (1918) or Fedosyutkin and Nainys (1993) (see p. 81) depending on which was deemed to fit best. All other aspects of the nose were left to the normalization of the average face shape, so long as the bridge of the nose aligned with the nasal bones etc.

Philtrum

Philtrum placement was left to computer normalization of the average face shape to the skull but was subjectively adjusted if it seemed unreasonable.

Mouth

Cheilion positions (landmarks 73 & 79) were determined according to tested guidelines presented in this thesis (see p. 63 to 79) to establish mouth width. Stomion position was determined according to George (1987), i.e., stomion is located ¼ way up the central incisors for males and 1/3 way up the central incisors for females, since his guidelines have been based on direct anatomical relationships while other guidelines suggested for determining this feature have not been reported with associated evidence. Lip height was determined in accordance with the subjective guideline that it is equal to the enamel height of the upper and lower incisors (Gatliff, 1984). Lip shape was subjectively determined although being based on the standardized average face shape.

Folds and lines of the face

The mental fissure was subjectively placed above the mental sulcus. The mentolabial groove was subjectively placed above mentolabial sulcus as indicated by George (1993). The inferior orbital groove was subjectively oriented consistent with general anatomy as were the delineation contours indicating cheek prominence. The nasolabial groove was located using the standard average faces as guides. The epicanthal folds (or superior orbital grooves) were left to the computer normalization of the average face shape, but altered by small amounts if contour lines overlaped. Other lines and grooves (e.g., vertical glabella lines, transverse nasal lines, inferior orbital grooves, oromental grooves etc. of the face (see George and Singer (1993)) were dependent upon the anatomy of the average faces and other delineated face shapes.

Ears

Height of ears (landmarks 181,186 & 187, 192) were determined according to general guidelines of Farkas (1987), taking into account magnification effects caused by differing feature distances to the camera. However, since Farkas does not report how much larger most people's ears are than their noses, subjective interpretation was used. All other ear anatomy was determined using the standard average faces as guides and ensuring that contour lines did not overlap.

Hairline

Hairline was subjectively estimated.

Neck

Neck outline and sternocleidomastoid muscle outline were subjectively estimated.

Profile Facial Approximation:

General face outline

The 2D average face shape was adapted to the skull in frontal view using landmarks otobasion inferius (landmark 1), exocanthion (landmark 2), and part of the upper lip border (landmark 3) as references (see Fig. 33 or Appendix 7). Adjustments were then made to the normalized 2D face shape delineations, as required, to accommodate the underlying skull and average soft tissue depths (Fig. 53 to 58). Again, relaxed average faces were used since skull casts were mounted with a freeway space. All landmarks described below can be located in Figure 33 and Appendix 7.

Eyes

Anterior eye projection (landmark 15) was determined using exophthalmometry values as presented in this thesis (p. 109 to 112). Palpebral fissure shape was based on the standard average profile, although it was subjectively determined.

Eyebrows

The eyebrows were based on the standard average delineation profiles.

Nose

Position of pronasale (landmark 107) was determined according to new methods described in this thesis (p. 91 to 96). Alar shape was subjectively determined and positioned using the standardized average face profile and observations of George (1987). Nasal bridge form was determined according to skeletal face profiles (Robinson *et al.*, 1986) and nasal bone form. Shape of tip of nose and columella angle were entirely subjective.

Mouth

Lip shapes were subjectively determined but based on the standardized average shape. Cheilion (landmark 20) placement was subjective again being placed based on the standardized average shapes, which indicated it was close to the same hosizontal position as the alar curvature point, well forward of cheilion positions as indicated by George (1987).

Folds and lines of the face

The mentolabial groove was subjectively placed above mentolabial sulcus as indicated by George (1993). The inferior orbital groove was subjectively placed consistent with general anatomy as was the delineation contour indicating cheek prominence. The nasolabial groove was located using the standard average faces as guides. The epicanthal folds (or superior orbital grooves) were left to the computer normalization of the average face shape. Other lines and grooves (e.g., vertical glabella lines, transverse nasal lines, inferior orbital grooves, oromental grooves etc. of the face (see George and Singer, 1993) were dependent upon the anatomy of the average faces and other delineated face shapes.

Ears

The ears were positioned about 5mm posterior to the external auditory meatus, and about 5mm superiorly, according to guidelines by Welcker (as cited by Stewart (1979b)). Ear height (landmarks 87 & 82) was determined as indicated above for frontal views. Ear angulation was set at about 15 degrees posteriorly inclined from the vertical (Gatliff and Taylor, 2001) but ensuring that it was less than the nose angulation as suggested by Farkas (1987) and as indicated by the standardized average faces. Other shape morphology of the ear was modeled based on the standardized average faces.

Jaw outline

Jaw line was subjectively placed slightly inferiorly to the lower border of the mandible.

Hairline

Hairline was subjectively estimated.

Neck

Neck outline and sternocleidomastoid muscle outline were subjectively estimated. However, sternocleidomastoid muscle direction was antero-inferior from the mastoid process, unlike directions of Neave for it to run directly inferiorly (R. Neave, 2002, personal communication) (see Fig. 53).

Figure 54 to 59 show superimpositions of the delineation maps over the skull, the facial approximations (i.e., the warped average faces), and superimpositions of the facial approximations over the skull, in both frontal and profile views. Figure 60 shows the final facial approximations as presented to assessors.

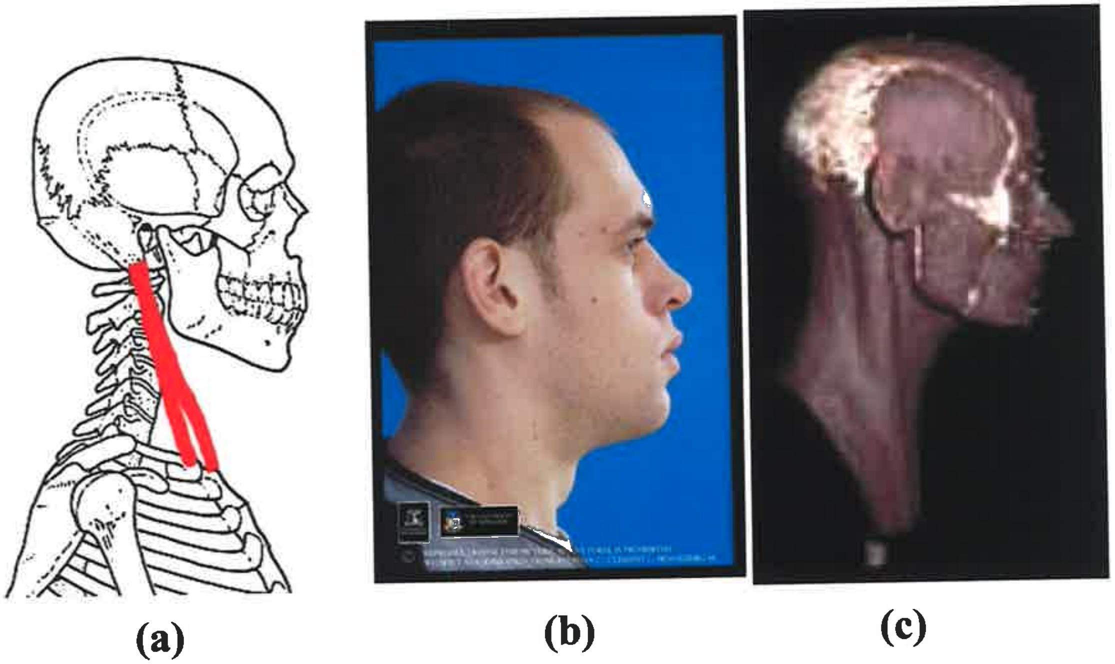


Figure 53: Sternocleidomastoid direction

(a) skeletal morphology and muscle relationship, adapted from Gray's Anatomy (Williams, 1995) p.427; **(b)** clear demonstration of relationship in (a) in a living subject; **(c)** image of facial approximation by C.S. Milner, R.A. Neave, and C.M. Wilkinson, illustrating the construction of inferior only directed sternocleidomastoid muscle; reproduced with permission from Milner *et al.* (2001) © Forensic Science Communications.

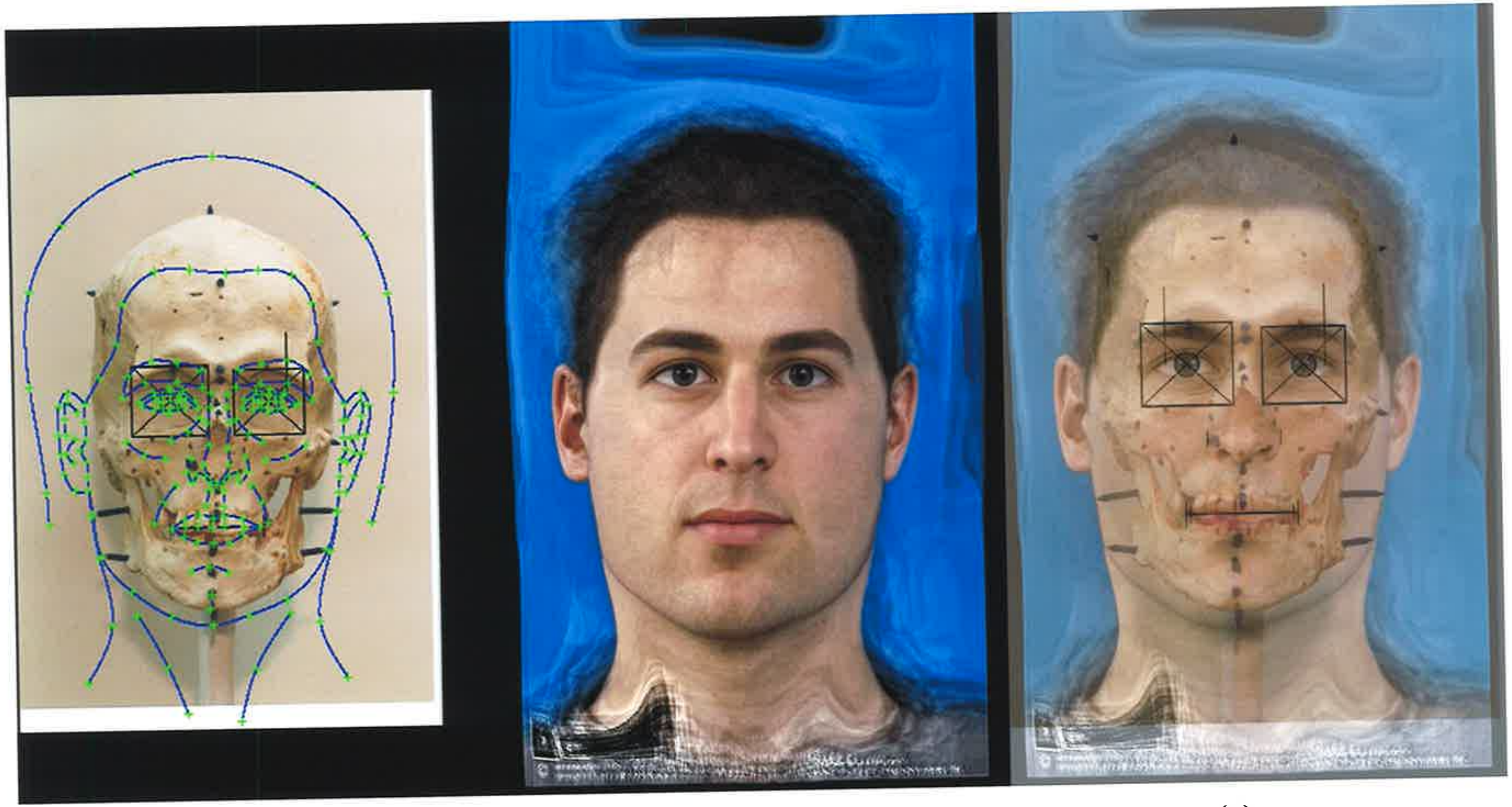


Figure 54: Frontal facial approximation of "Fred"
(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

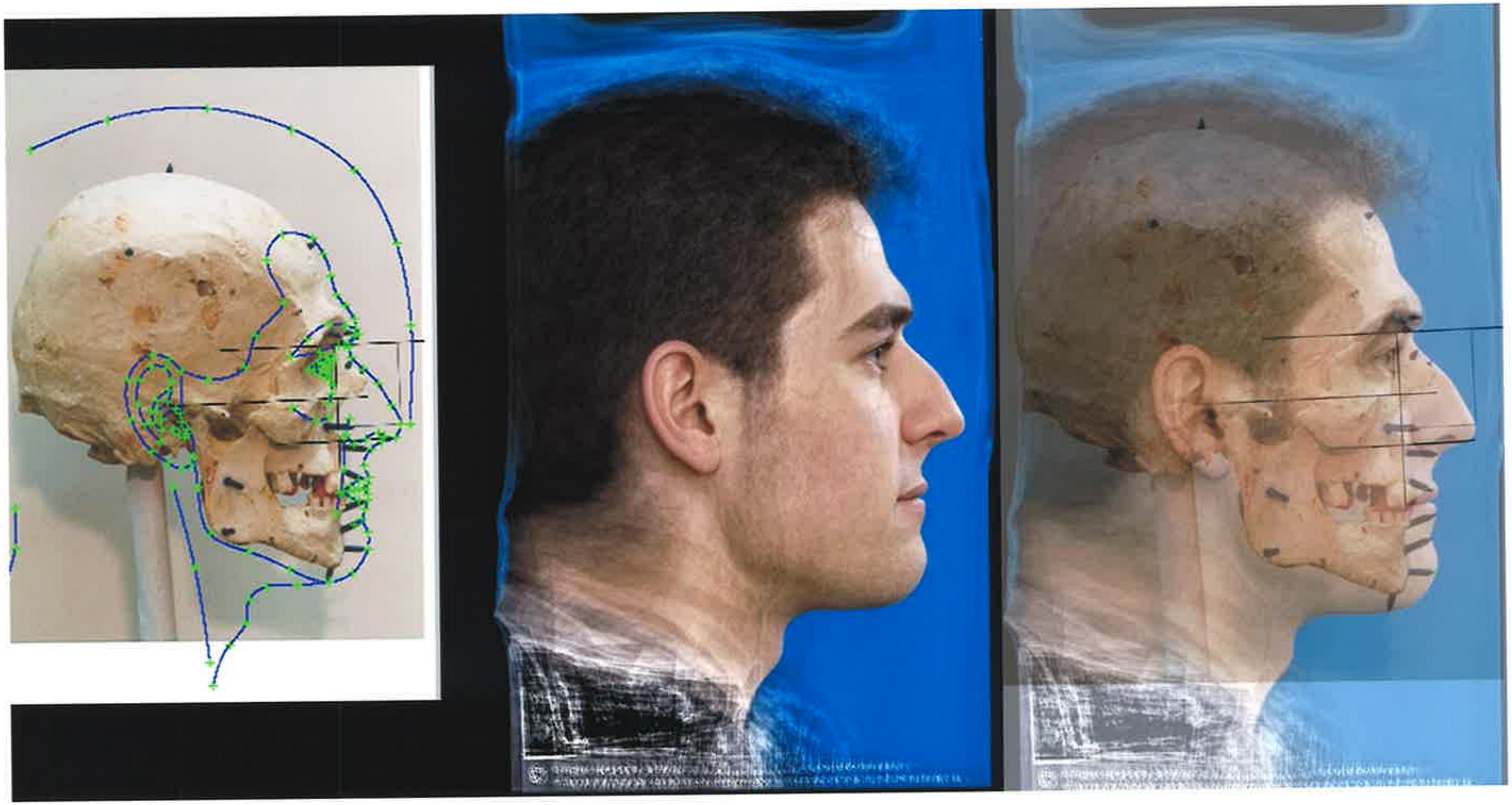


Figure 55: Profile facial approximation of "Fred"

(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

Figure 56: Frontal facial approximation of "Jane"

(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

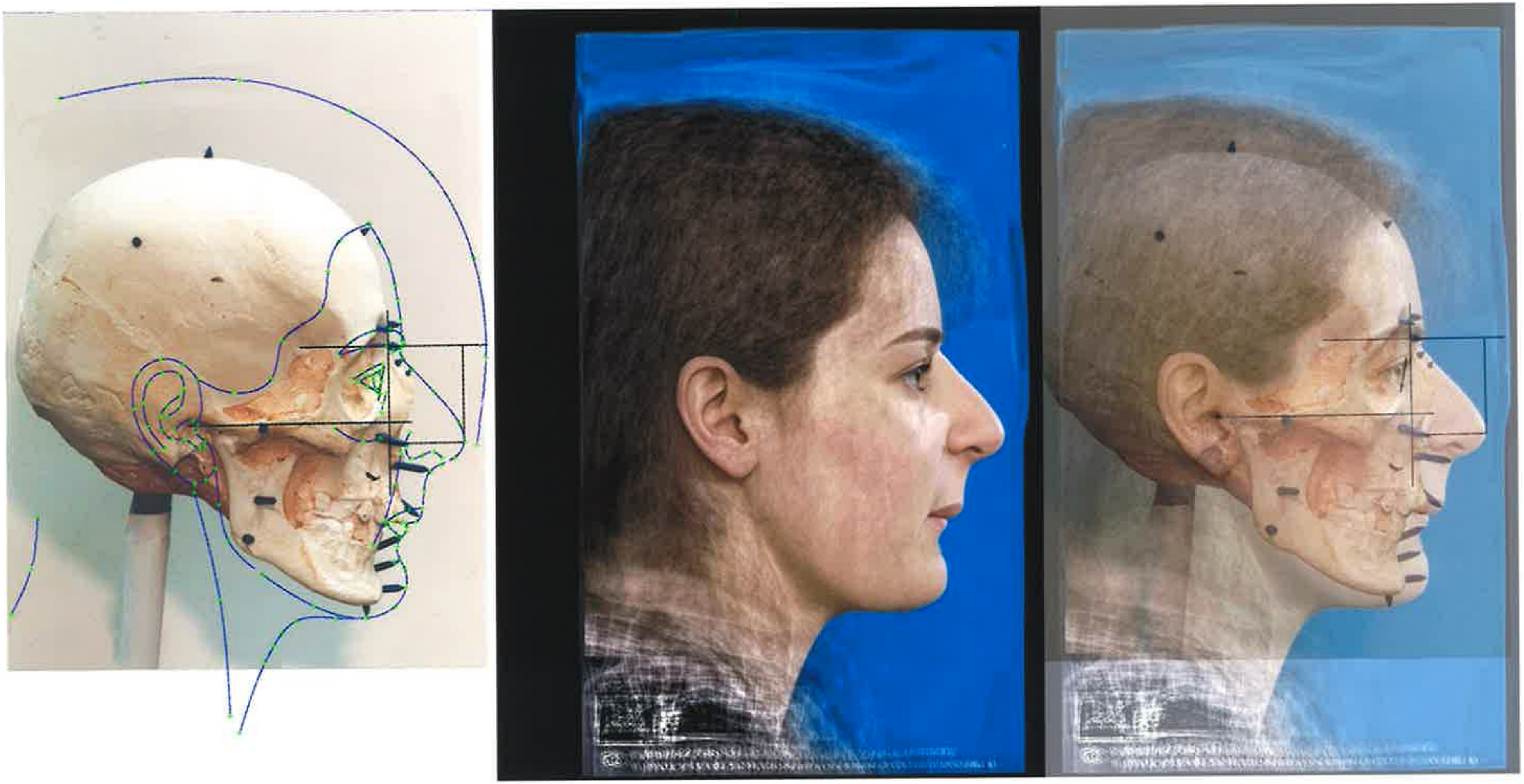


Figure 57: Profile facial approximation of "Jane"

(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

Figure 58: Frontal facial approximation of "Kate"

(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

Figure 59: Profile facial approximation of "Kate"

(a) superimposition of delineation map with skull. **(b)** average face warped to delineation map (facial approximation). **(c)** superimposition of skull with facial approximation.

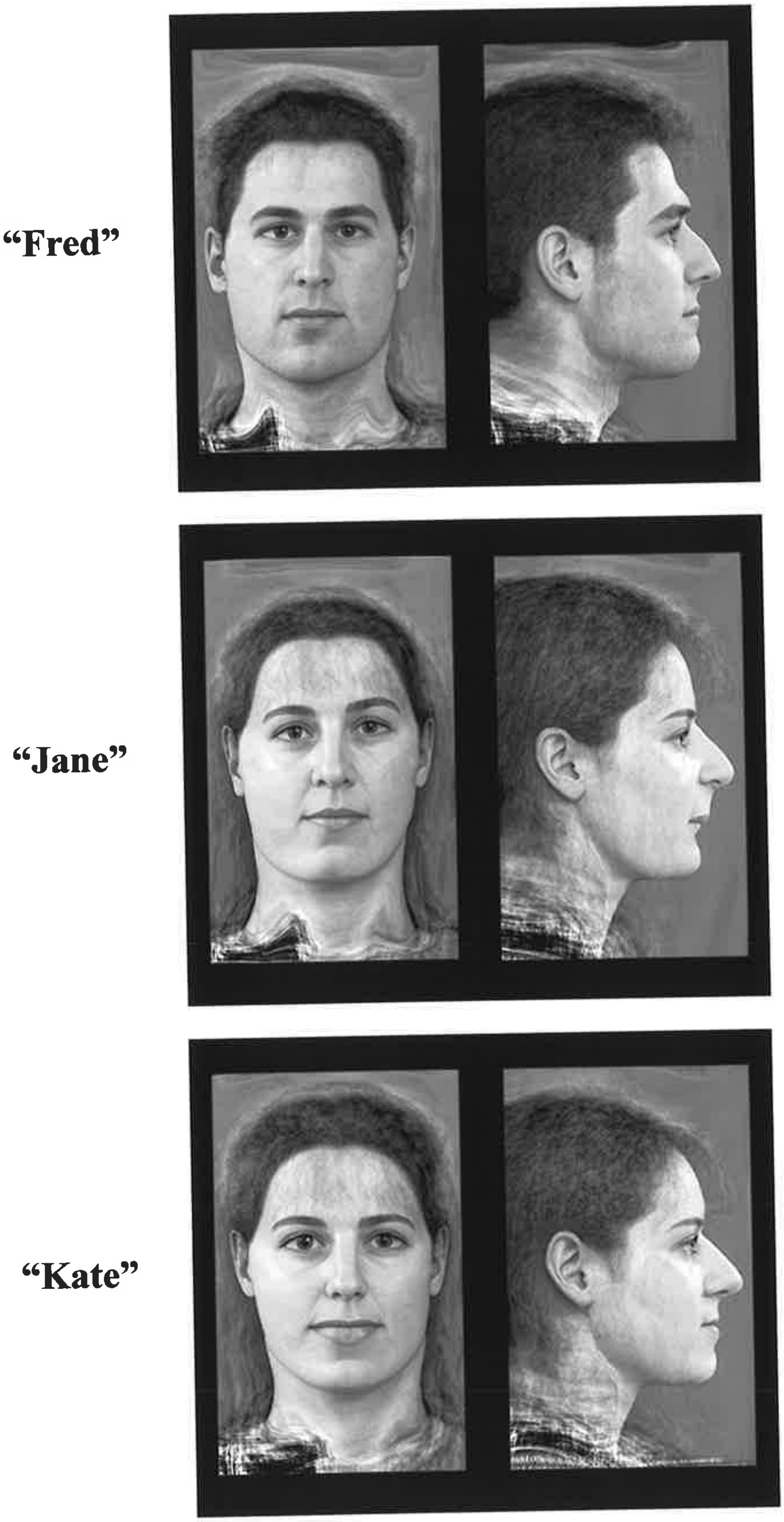


Figure 60: Facial approximations as presented to assessors

Testing Facial Approximation Recognition

Thirty two assessors (16 females, mean age 28 years, SD 13 years; 14 males, mean age 39 years, SD 14 years) participated in the recognition trials where attempts were made to identify target individuals, displayed in face pools, from facial approximations. Assessors were presented with facial approximations (with both front and profile views) one at a time, in a predetermined random order, and they attempted to identify from the corresponding face pools for each, who the target individual was. Face pools included the same 10 individuals in the same order as used by Stephan and Henneberg (2001). Assessors were told that target individuals may or may not be in all face pools. When target faces were present subjects had a 50 % chance of deciding if the target face was actually in the face pool and a 10% chance of correctly identifying the correct face. Therefore, assessors had a 5% chance (50% x 10% = 5%) of correctly identifying the target individual when they were present. Although participants were not aware of the fact, all face pools included the target individual. Since one of the participants recognized one of the non-target faces in one of the face pools his response for this trial was not included in the study. This gave a sample size of 32 individuals for "Fred", 31 individuals for "Jane", and 32 individuals for "Kate".

For each facial approximation assessors also indicated what their confidence was that they had made the correct decision (i.e., identified the right face or correctly decided that the target face of the facial approximation was not in the face pool), and if a face was identified, what s/he felt the resemblance between the facial approximation to the face identified was. The same scales were used as presented on p. 165 and 166, with confidence level being indicated by: not confident, slightly confident, fairly confident, or confident; and resemblance being indicated on a scale from 0 to 10. Data were analyzed using Fishers exact tests (JMP® 4.0), and unequal variance t-tests (Microsoft Excel® 2000).

Results

None of the three facial approximations constructed were correctly identified above chance rates. Out of a total of 95 identification scenarios there were 4 true positive identifications (4%), 66 false positive identifications (69%), and 25 instances where no identification could be made (26%). Table 25 summaries these results. Target individuals for facial approximations of “Fred” and “Kate” were never identified while that of “Jane” was identified 4 times (13%), but this rate was not above that expected by chance (Fisher’s exact test, p>0.05). The only face identified above chance rates was a non-target face, photo number 1 in the face pool for “Fred” (Fisher’s exact test, p<0.01). This face was identified 13 out of 32 times (41%).

Table 25: Facial approximation recognition responses
Target face identifications for each facial approximation are shown in bold type.

Facial Approximation Trial	**Target Face Identification**											**Total**
	1	2	3	4	5	6	7	8	9	10	No ID	
"Fred"	13	**0**	0	0	7	1	0	1	0	0	10	32
"Jane"	**4**	3	5	0	4	5	1	0	0	2	7	31
"Kate"	3	7	1	**0**	1	0	3	7	2	0	8	32

Average resemblance ratings of true and false positive identifications did not differ at statistically significant levels (unequal variance t-test, p>0.05). If anything there was a trend for false positive identifications to receive higher resemblance ratings than true positive identifications (Fig. 61). Resemblance ratings also displayed trends to increase with increasing confidence levels of assessors (Fig. 62).

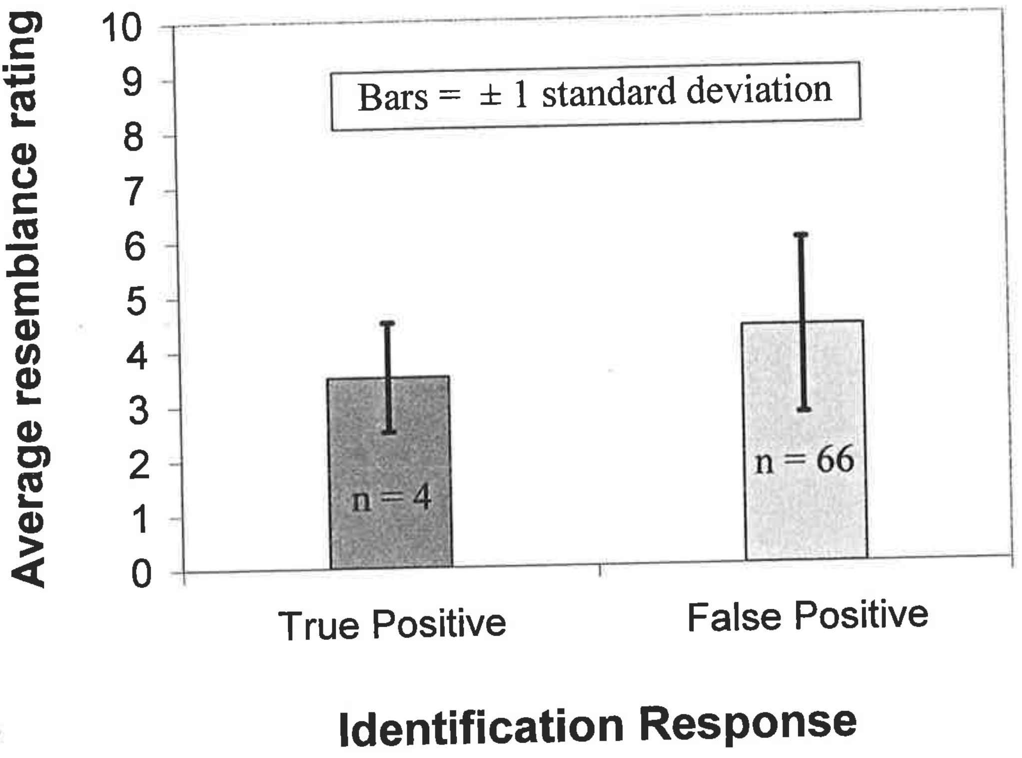


Figure 61: Resemblance ratings for facial approximation identifications

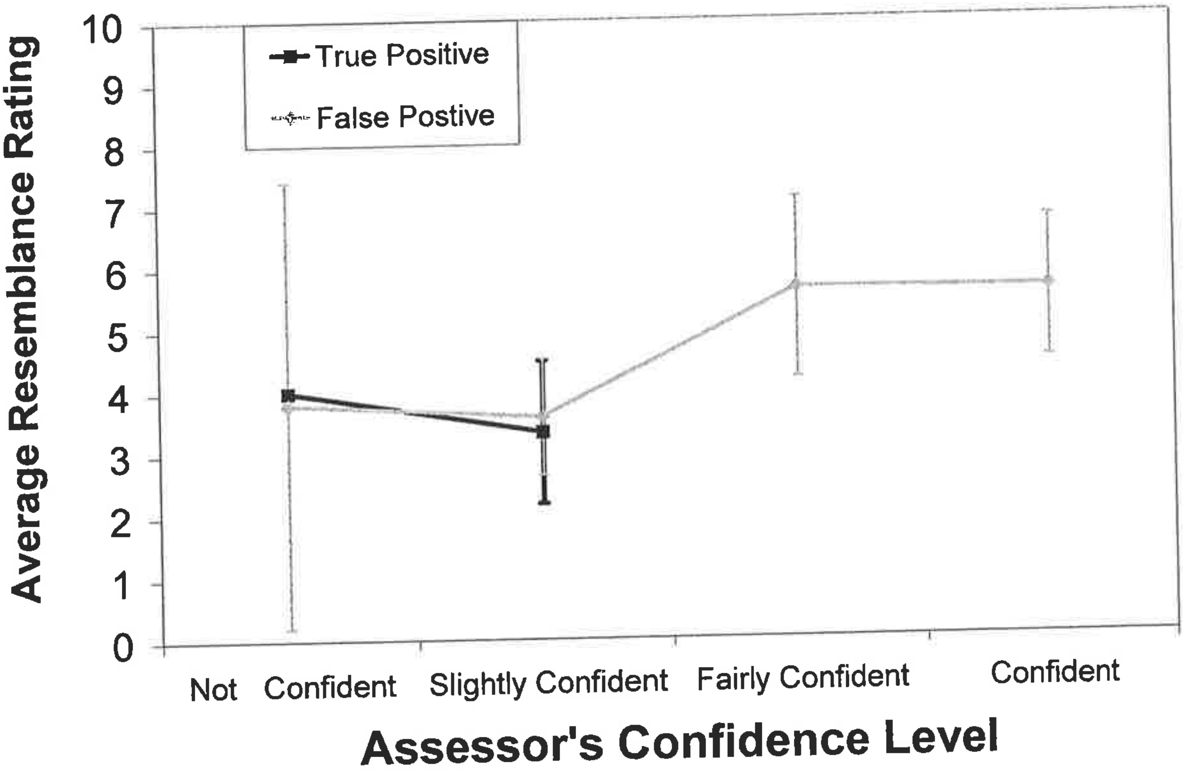


Figure 62: Relationship between resemblance ratings of facial approximation identifications and assessors' confidence levels

Discussion

The lack of above chance true positive identifications of facial approximations even when they are constructed from less subjective methods, i.e., empirically tested and improved soft tissue prediction guidelines and average faces, demonstrates that facial approximation methods are inaccurate. It was expected here that true positive recognition rates would be less than ~43% since exact face shapes could hardly be expected to be replicated (as indicated previously), however, actual true positive recognition rates were much lower – in two cases none were made.

These results are not surprising since exact face shapes can hardly be expected to be created from the skull using the few tested guidelines that exist, many of which have only been described here for the first time. Additionally, since traditional facial approximation methods are subjective and use inaccurate soft tissue prediction guidelines, it seems that these facial approximations are likely to be even less successful. However, the results of this study compare closely to those reported by Stephan and Henneberg (2001) even though they used more subjective methods. For example, true positive identifications were found 4% of the time here, and 8% of the time by Stephan and Henneberg (2001), false positive identifications were 69% and 67% respectively and no identification instances were 26% and 25% respectively. It, therefore, seems that facial approximation recognition rates may be at base levels, even when improved techniques described here are used. This suggests much work is still needed if facial approximation is to become even a partially reliable technique that achieves true positive identifications purposefully (not due to chance) at least some of the time. This study reinforces conclusions of Stephan and Henneberg (2001) that facial approximations rarely result in true positive identifications.

Results of this study also replicate those found earlier in this thesis that resemblance ratings are not reliable indicators of the recognition ability of facial approximations and hence their accuracy. It has been suggested, apparently in an attempt to justify the continued use of resemblance ratings, that the finding that facial approximations sometimes (perhaps often) bear greater resemblance to non-target individuals and that they are identified instead of target individuals is logical and not surprising (Wilkinson and Whittaker, 2002). However, the argument that non-target individuals often represent facial approximations is precisely one of the reasons for why resemblance ratings should not be used to gauge a facial approximations accuracy, which ultimately rests on its "recognizabiltiy".

It has also been suggested that previous findings of the author that facial approximation recognition is low and infrequent (Stephan and Henneberg, 2001) may have resulted because of a lack of practitioner experience with techniques (Wilkinson and Whittaker, 2002). However, since this study used less subjective methods and closely adhered to new improved and empirically demonstrated guidelines and used computer generated average human face morphology, this study clearly demonstrates that facial approximation results in low and infrequent identifications of target individuals irrespective of practitioner experience and the "artistic quality" of the facial approximations.

Figures 63 and 64 present the facial approximations constructed here using more objective methods and those of Stephan and Henneberg (2001) so they can be compared. The figure does not present the facial approximations according to scale but rather the images are presented to allow holistic comparisons to be made. The "artistic quality", perhaps better described as how real or lifelike the facial approximations are, is definitely fairly low in all facial approximations presented in Figures 63 and 64. As mentioned in the introduction this is primarly due to the fact that attempts were made to restrict subjective inputs into the facial approximations, since any such characteristics added are purely “intuitive”. As a result two-dimensional drawings and FACE generated approximations have minimal tone and highlights since this requires subjective estimation. Also it can be seen that few wrinkles and lines or grooves of the face are included since these cannot be determined from the skull. This may contribute to the low recognition of these facial approximations but this again demonstrates that if detailed subjective information is included and the face is recognized correctly that it is likely a result of chance. Increasing the detail of facial approximations to make them appear more realistic and of higher “artisitic quality” is definitely favourable but this should also be done by employing methods which have been empirically demonstrated to be reliable and accurate. By using average face colour and texture more facial detail can be objectively included. Figure 65 and 66f demonstrate the true artistic ability of the author since these heads have been constructed with little restriction.

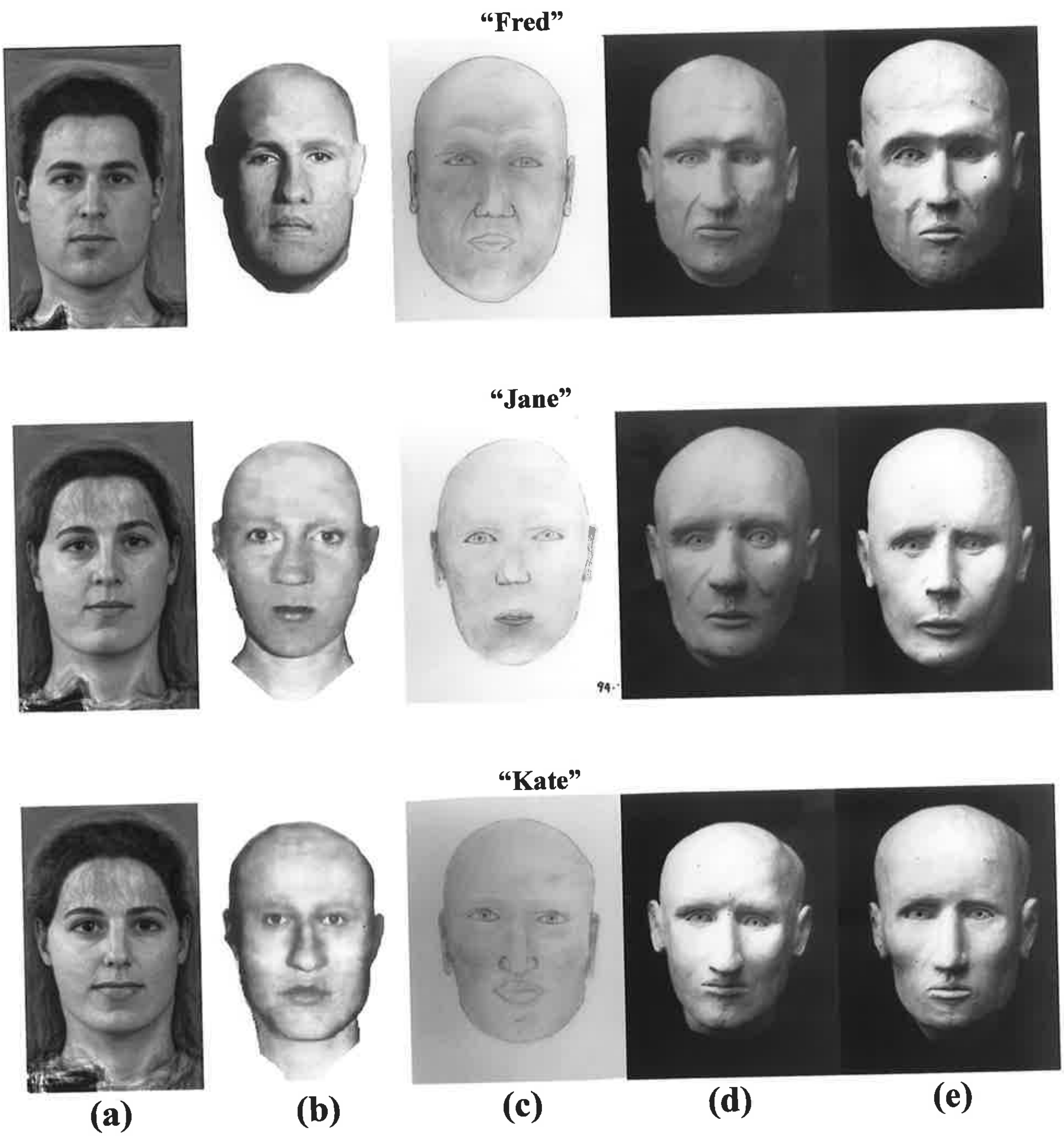
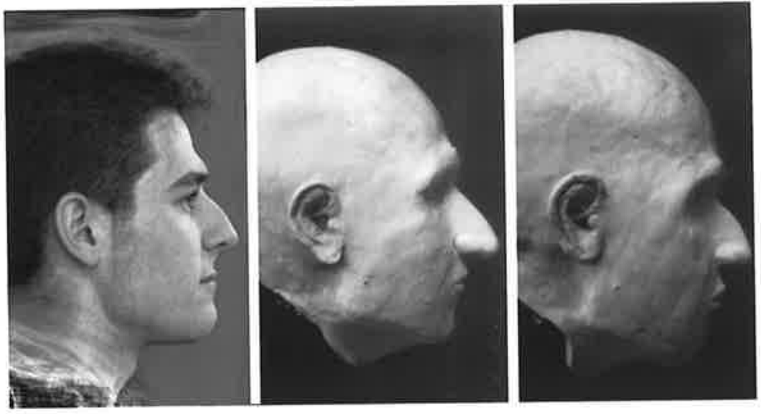


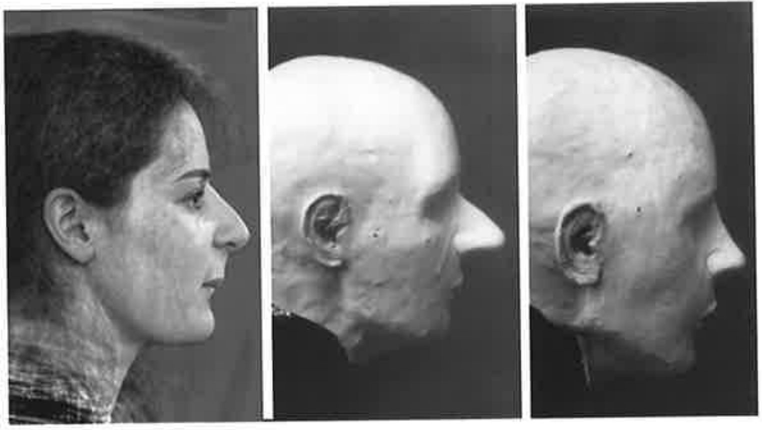
Figure 63: Comparison of frontal facial approximations for "Fred", "Jane" and "Kate"

(a) the facial approximations constructed here; **(b)** to **(e)** those constructed by Stephan and Henneberg (2001). Images are not to scale. See above for methods of **(a)**. For precise soft tissue predictions for **(b)** to **(e)** see Stephan and Henneberg (2001) p. 433-435. **(b)** 2D Face Assisted (Photo-Fit Computer) method. **(c)** 2D drawing method. **(d)** 3D combination method. **(e)** 3D American method.

"Fred"



"Jane"



"Kate"

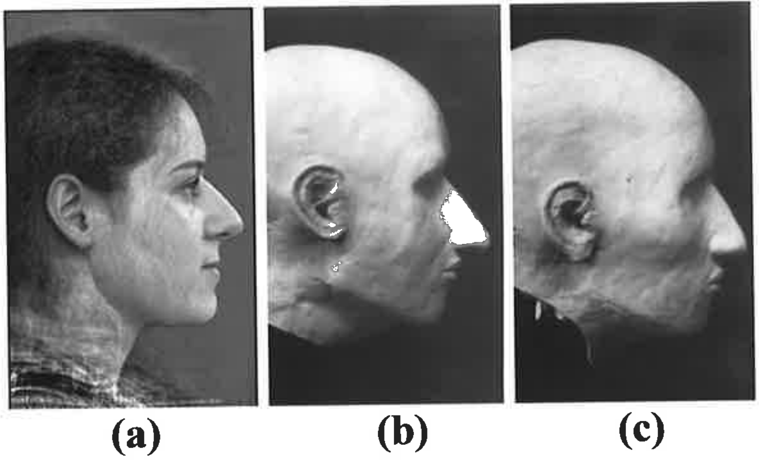


Figure 64: Comparison of profile facial approximations for "Fred", "Jane" and "Kate"

(a) the facial approximations constructed here; **(b)** and **(c)** those constructed by Stephan and Henneberg (2001). Images are not to scale. See above for methods of **(a)**. For precise soft tissue predictions for **(b)** and **(c)** see Stephan and Henneberg (2001) p. 433-435. **(b)** 3D combination method. **(c)** 3D American method.

(a) **(b)**

Figure 65: Artistic facial construction by the author

(a) Portrait of Francois Vidocq; and **(b)** Clay sculpture based on portrait. Made for the Australian and New Zealand forensic Science Society for display at the 14th International Symposium on the Forensic Sciences, Adelaide, Australia, 1998.

Each series of facial approximations constructed using different methods on each skull (Fig. 63 and 64) appears to share a rather "high" degree of similarity with a consistent "theme" running between many of the faces. Despite this there are also many differences, some due to methodological approaches. Particularly prominent are the large noses that result when using the Gerasimov (1971) (tangents following general directions of nasal bones) guideline for nose projection in the 3D combination approach (Fig. 64). The shape of the nose in profile also differs between facial approximations of the same skull as a result of different techniques producing different nose projections, which need to be accomodated to create a "balanced" or realistic appearing face. The 2D FACE assisted method of "Kate" certainly seems to have produced a face that is too long, however, the same skull picture was used for the drawing method. A possible explanation for this observation is that the nose of the FACE assisted approximation seems prominent and also a rather projecting chin has also been used, making the face appear longer. As previously described by Stephan and Henneberg (2001), this is a limitation of that technique since limited photofit features are available to choose from in the construction of these facial approximations. A similar scenario can be seen for “Fred” where his jaw line and cheeks appear to be “fuller” than in the other facial approximations. Despite differences between facial approximations recognition tests showed that facial approximations of “Fred” frequently resulted in incorrect indentifications the face pool photo 1 both in this study and that reported by Stephan and Henneberg (2001) for 2D methods. Likewise, face pool photo 6 for “Jane” was recognized frequently here as it was when using 2D methods were used by Stephan and Henneberg (2001). Results obtained for the facial approximation of “Kate” did not display consistent trends between the studies however. This seems to indicate that there is consistency, at least between facial approximations created by the same person, using similar methods (in this case two-dimensional ones) on the same skull.

Despite the lack of success in facial approximation recognition in this experiment, the adaptation of the software capable of warping average colour and shape to particular shapes seems beneficial to the facial approximation process since it dramatically limits subjective interpretation. The methods described above are also useful because they are easy to follow, logical and clearly indicate what is being subjectively estimated. Although the use of standardized average faces as models for the basis of determining unknown face features is favourable in contrast to purely subjective methods, some subjectivity is still involved as clearly indicated by the methods.

The techniques used above have several limitations: (i) few systematically tested guidelines exist for predicting the soft tissues of the face; (ii) systematically tested guidelines that do exist are averages; (iii) face morphology is represented in 2D and therefore much face shape is represented by colour changes, e.g., cheek shape; (iv) colour and texture information of target individuals outside the ages of the samples used for generating the average faces will not be well represented by the averages. Almost all these limitations could be reduced with future work. For example, more systematically tested soft tissue prediction guidelines could be determined; soft tissue prediction guidelines taking into account other feature relationships could be constructed using regression; average faces could be generated in three dimensions; and average faces could be generated for many age subgroups of the human population.

In two-dimensional (2D) facial approximations the use of average soft tissue depths is also limited because only those perpendicular to the line of view are useful. Soft tissue depths that are not perpendicular to the line of view are not useful since they cannot be used to precisely represent the soft tissue depth, which must be achieved by toning, e.g., depth at pogonion in frontal view. Hence, using these depths in 2D facial approximations is illogical for they serve little purpose but to make the technique appear more precise than it is, which again overemphasizes its rigor to lay individuals. Here it was found that only 10 individual soft tissue depths were relevant for frontal 2D approximations and 12 soft tissue depth in profile, where as other authors unrealistically use many more, e.g., Taylor (2001a) uses about 29 depths in frontal views 21 in profile views.

Exocanthions and endocanthions were predicted here according to the commonly cited methods of Krogman (cited in Caldwell, 1986), which appear to be consistent with reported observations of others (Angel, 1978) even though these studies do not provide any empirical justification for these guidelines (Table 26). However, Yoshino and Seta (2000) also present their own different methods for determining the lateral canthus, i.e., that the exocanthion fall on or just medial to the lateral orbit. Since conducting the experiment reported above further literature searches have revealed evidence that shows all three of the above guidelines to be incorrect. A study by Rosenstein and colleagues (2000) determined the distance the lateral canthus falls away from the lateral orbital wall, relative to the coronal plane (Rosenstein, 2003, personal communication) in 21 individuals. They found the average distance to be 7.5

mm (SD = 2.8mm, range: 2-12mm), a value between two- and seven-times greater than that indicated by traditional facial approximation guidelines. Furthermore a study by Gioia and colleagues (1987) in 1987 reported bilateral dissections of 8 human cadavers where they were able to measure the long axis of the lateral canthal tendon and calculate its average length (n=16). The value they found was 10.6 mm, with a standard deviation of 0.9mm. Since the lateral canthal ligament attaches on average 1.5mm behind the orbital margin (Gioia *et al.*, 1987), the length of the lateral canthal tendon reported by Gioia and colleagues (1987) appears to corroborate findings by Rosenstein and colleagues (2000), as together they indicate that the lateral canthus falls about 6mm anterior to the lateral orbital margin - a value that seems reasonable given that the cornea projects about 16mm anterior to the lateral orbital rim (as reviewed in this thesis). Therefore, while the results of Rosenstein and colleagues' study (2000) suggest that Krogman's (cited in Caldwell, 1986), Angel's (1978) and even Yoshino and Seta's (2000) reports are possible in some instances, on average they will tend to grossly under predict the distance the lateral canthus falls away from the lateral orbit in the majority of individuals. This suggests that the lateral canthi positions of the facial approximations used in this experiment are incorrect, as they are in any previously constructed facial approximation. This observation once again demonstrates that untested facial approximation guidelines based on observational evidence are misleading and further reinforces the need for all such subjective guidelines in facial approximation to be empirically evaluated. It therefore seems that facial approximations should be constructed in the future using the values of Rosenstein and colleagues (2000) until other larger sampled studies have been conducted.

Table 26: Reported values of lateral canthal distance from the lateral wall of the orbit

	Untested Reports		Tested Reports
	Angel (1978)	Krogman (cited in Caldwell 1986)	Rosenstein et al. (2000)
Average distance of lateral canthus from lateral orbital wall	3 to 4	5	7.5
SD	-	-	2.8
sample size (n = number of cadaver specimens)	-	-	**21**

It also seems worth noting that many assesors after identifying who they thought were the target faces, were told who the target individuals actually were, they would often say things like "oh, I was thinking it might be him/her, they were going to be my second guess". One wonders if this really would be the case, only testing it would resolve the matter. However, it maybe that because the facial approximations often bear some type of resemblance to almost all the faces and this could be said for any. Many assessors also commented that the longer they looked at the faces in the face pools the more similarities they saw between the facial approximation and *all* of the faces in the face pool. As is expected since average methods are used in the facial approximation process, this seems to indicate that the facial approximations do not display any distinctive individualising features. Perhaps the Class II occlusion of "Jane" counted as a more distinctive feature, which resulted in her taget face being identified more often than the other two facial approximations. This highlights the limitations of using average approaches to facial approximation and the importance of drawing attention to distinctive features of the face as has been previously suggested (Taylor, 2001a; Ubelaker, 1993).

General Discussion and Conclusions

The main findings of the studies presented here are: (i) that many of the traditional guidelines used to predict major soft tissue features of the face from the skull could be improved (like those for determining eyeball projection, pronasale position, mouth width and superciliare position); (ii) resemblance ratings do not measure a facial approximations accuracy, that is, ability to be recognized; (iii) the highest recognition rates of two-dimensional facial approximations are ever likely to achieve, i.e., when exact face shapes are represented, are about 43% on average; and (iv) that facial approximations constructed from objective methods using average face anatomy and updated soft tissue prediction guidelines do not produce frequent true positive recognitions, as measured by face pool comparisons.

Since "objective" methods of facial approximation were tested here by using empirically demonstrated soft tissue prediction guidelines and average human faces, the results clearly demonstrate that facial approximation methods do not currently produce faces that can be recognized above chance rates. The fact that all of the *commonly* used soft tissue prediction methods tested here were rather inaccurate and unreliable suggests that the many other commonly used but untested subjective guidelines in circulation could also be improved. However, even with the best improvement possible, so that exact face shapes can be recognized it seems facial approximation will not obtain recognition rates higher than ~43% on average and that in some cases individuals will not even be recognized correctly.

Previous facial approximation success can be explained through mechanisms other than facial recognition, like contextual information (Haglund and Reay, 1991) or chance (Stephan and Henneberg, 2001). It may also be that facial approximation acts to promote public interest in the case, which in turn generates extra leads through the mechanisms suggested above, without the facial approximation being specifically recognized during critical stages. It is important to note that people probably *think* they recognize the target individual from the facial approximation *after* an identification has been made. It therefore seems that despite the inaccuracy of the technique, facial approximation is useful to forensic work since it may generate public interest and act as a "vehicle" for identification. However, it should be recognized that successes due to specific and purposeful facial recognition are unlikely, with recognitions probably being made independently of the facial approximation. Stewart (1979b) p. 257, appears to have been correct when he states that "likeness *per se* [may have nothing] to do with [facial approximation recognition]….[the] methods are surrounded by an aura of glamour that may not always be warranted". Indeed participants in face recognition experiments found the task difficult and often reported that the longer they looked at the faces in the face pool the more facial approximation seems to resemble all of the faces. This highlights the fact that the faces built from skulls do not display characteristics that make these faces easily recognizable as their target individuals. While specific and purposeful facial recognitions may appear to be made in forensic cases, since family members etc. recognize the facial approximation of the victim, this may not be the case. People who are aware that a loved one is missing may be more likely to respond to calls for recognition from advertised facial approximations. Since victims cannot be of individuals who are known to be alive it is expected that identifications from individuals who have a missing acquaintance will have an increased likelihood of being correct.

While studies presented here show "beyond reasonable doubt" that traditional facial approximation methods are inaccurate and unlikely to result in specific and purposeful facial recognition, this finding does not indicate that facial approximation is not useful or harmful to identification attempts as has been suggested by others (e.g., Wilkinson and Mautner, 2003; Wilkinson *et al.*, 2002). This point needs to be emphasized since it has been repeatedly stated in the literature that these findings show that "facial approximation would be detrimental to any forensic identification case" (Wilkinson and Mautner, 2003;) p. 12, and (Wilkinson *et al.*, 2002) p.112. Such statements are not only incorrect but they misrepresent the conclusions drawn by the original authors. Any facial approximation even if recognized purely by chance will be useful for forensic identifiction. It is essential in forensic scenarios that every method possible be used in an attempted to identify skeletal remains, not only for the important task of easing family concern and even grieving, but also to assist in the apprehension of the offender/s (Brues, 1992). A logical procedure is to use those methods that are most reliable and accurate first and work down the list to the least accurate and least reliable methods if the former happen to be unsuccessful for any reason. Therefore, if more reliable methods cannot be used or have proven to be of little use in establishing an identification, but facial approximation methods can be used, then they definitely should be since the may, by chance or by purpose (however unlikely), help to establish an identification.

While results of studies presented here demonstrate that 2D facial approximation techniques do have potential to obtain personal recognition in a forensic setting, the inaccuracy of specific soft tissue prediction guidelines as found in the studies above suggests that much research in soft tissue prediction from the skull is needed if this recognition potential is to be realized. The research results found in this thesis strongly confirm suspicions by others that the reliability, accuracy, and scientific rigor of traditional facial approximation methods have been overstated considerably in the past. Consequently, the extremely small portion of systematic information previously used in the process of facial approximation does not seem to justify the method being referred to as a blend of "science and art" nor does it seem to justify expectations that the method works. Only in the future when the majority of the method comes to rest on tested (but even if inaccurate) guidelines can the technique, as a whole, be considered "scientific". If the ability of facial approximation methods to achieve high rates of purposeful, specific and reliable facial identification is to be demonstrated, it must be done under experimentally controlled conditions since forensic casework success may be influenced by other factors as mentioned above.

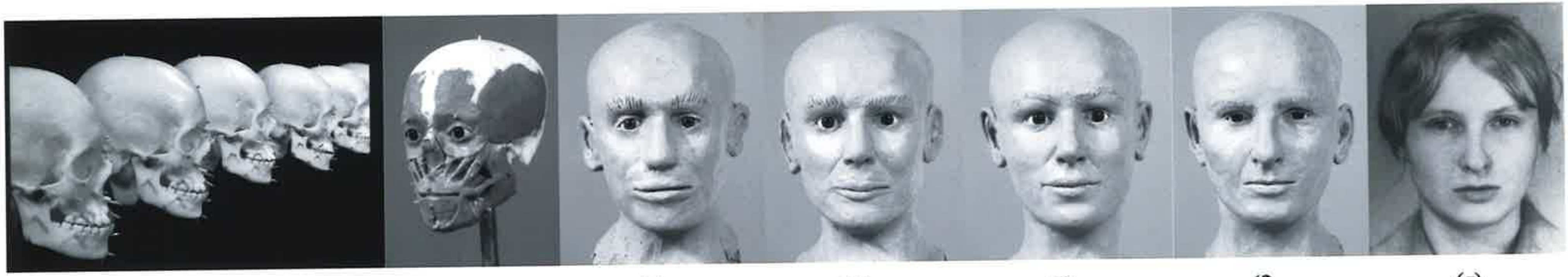
The formulation of scientific soft tissue prediction guidelines has the advantage that prediction errors may be correlated with recognition rates, allowing prediction errors to be used as a quantitative measure of the "recognizability" of facial approximations. Such a measure would be beneficial since it could be generated without the use of timely face pool comparison experiments, which require the identification of the target individual *before* the accuracy of the facial approximation can be determined.

With further systematic studies of the relationship of the soft tissues to the skull facial approximation is likely to come to include many technical soft tissue prediction guidelines. The regression equations determined in this thesis are examples of this, as are equations presented by Simpson and Henneberg (2002) in the literature that attempt to individualize average soft tissue measurements based on skull sizes. It seems probable that this will have ramifications for the way facial approximation is approached because some measurements are not easily taken from skulls directly (like nasal bone angle which may require planes to travel through the nasal bones if they are medially convex and rhinion is used as a landmark). Consequently, three-dimensional computer techniques will probably become the method of choice because they enable: indirect measurements to be made; the calculation of many variables in a very short time; and the generation of facial approximations in three dimensions. Therefore, there appears to be a need for facial approximation practitioners to move away from traditional three- and two-dimensional techniques and for three-dimensional computer techniques to be developed. Particularly disadvantaged are two-dimensional methods whether drawn, or computer generated since much of the average soft tissue depth information is of little use, i.e., only depths that run perpendicular to the field of view can be objectively used.

Given the current inaccuracy of facial approximation techniques it seems that employing a "specially trained practitioner" to do them is of little use since the process is highly subjective and any other sex, age, and population specific face may achieve similar results. However, public interest in the technique may decrease if the public becomes aware that facial approximation techniques are little more than "guesses". It therefore seems important that efforts are made to dramatically increase the accuracy of the specific soft tissue prediction guidelines so that the public remains interested in the technique and it continues to hold its bewilderment factor (e.g., the "Wow, can you really do that!" effect) and act as a "vechile" for identification.

Facial approximation is an extraordinary technique if it actually *works*, i.e., achieves its goal of specific and purposeful facial recognition of the target individual. This thesis demonstrates that some of the underlying soft tissue prediction methods used in the facial approximation process to determine major face features are not as accurate as some have thought. Even when improved guidelines were used, it was found that in the majority of cases facial approximation is unlikely to achieve specific and purposeful facial recognition of the target individual. However, this thesis also demonstrates that facial approximation has the potential to achieve its goal and generate above chance true positive recognitions. With more work in the future, facial approximation may come to realize specific and purposeful facial recognition of target individuals a hence become a more useful forensic technique.

The main priority for future research appears to be empirical determination of other soft- and hard-tissue relationships of the face for this would increase the accuracy of techniques. Since the face is likely to be influenced by evolutionary factors, studies elucidating the nature of any recent selection forces and their affects would be useful. The determination of standardized average soft tissue depth is also another priority since there are many tables proposed by many authors using many different techniques and there is little direction which measurements practitioners should be using. Furthermore, many soft tissue depths are reported by race and sex, which seems to be of little practitcal use in facial approximation since the differences in average measures between the groups are often less than 1 to 2mm and variation ranges between groups overlap considerably (Garlie and Saunders, 1999; Wilkinson, 2002; Williamson *et al.*, 2002). It would also be useful to determine what effect practitioner's dexterity (or "artistic quality" of constructed facial approximations) actually has on facial approximation recognizability (see Fig. 66) and whether or not face pools or sequential lineups are the best method for assessing facial approximation recognition. Additionally, it would be useful to test the recognizability of facial approximations that have been identified correctly in forensic cases using unfamiliar identification scenarios and face arrays since this would determine if those facial approximations were actually identified because they were recognizable or if some other factors played a role in their identification. With further research, the technique of facial approximation will be better understood and employed, increasing its usefulness in forensic applications. This knowledge will not only be useful to facial approximation but will also be highly relevant to skull-face superimposition techniques since they also rely on the knowledge of the soft to hard tissue relationships.



(a) (b) (c) (d) (e) (f) (g)

Figure 66: Facial approximations made by the author for a future study to test the influence of "artistic" quality on facial approximation recognition

Identical skull casts were produced according to methods of R. Taylor (a) and identical muscle buildups were conducted on each of the skull casts (b). Four facial approximations were then constructed by the author, of differing "artistic quality": (c) low artistic quality; (d) medium artistic quality; (e) normal artistic quality given rules of facial approximation; (f) high artistic quality - visage produced over skull cast with reference to target individual; (g) target individual (image courtesy of R. Helmer). Note that this study is being conducted together with R. Taylor who has also constructed similar facial approximations to test and that the casts used above were created by a team effort. Each facial approximation was constructed within a 10 hr time limit.

References

Alley TR, and Cunningham MR. 1991. Averaged faces are attractive, but very attractive faces are not average. Psychol Sci 2:123-125.

Angel JL. 1978. Restoration of head and face for identification. Paper preseneted at the 13th Annual meeting of the American Academy of Forensic Sciences, St. Louis, Missouri.

Aulsebrook WA, Becker PJ, and Iscan MY. 1996. Facial soft-tissue thickness in the adult male zulu. Forensic Sci Int 79:83-102.

Barretto RL, and Mathog RH. 1999. Orbital measurement in black and white populations. Laryngoscope 109:1051-1054.

Battagel JM. 1990. The relationship between hard and soft tissue changes following treatment of class II division malocclusions using Edgewize and Frankel appliance techniques. Eur J Orthod 12:154-165.

Behrents RG. 1985. Growth of the aging craniofacial skeleton: Craniofacial Growth Series. Ann Arbor.

Benson P, and Perrett D. 1992. Face to face with the perfect image. New Scientist 22nd February:26-29.

Benson PJ, and Perrett DI. 1991. Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. European Journal of Cognitive Psychology 3:105-135.

Bertelsen T. 1954. On 720 measurements with Lueddes exophthalmometer. Acta Opthalmol Scand 32:589-595.

Bogren HG, Franti CE, and Wilmarth SS. 1986. Normal variations of the position of the eye in the orbit. Ophthalmology 93:1072-7.

Bogren HG, Schermer MJ, Franti CE, Elfstrom G, and Tengroth B. 1976. Radiographic exophthalmometry. Transactions of the Academy of Ophthalmology and Otolarynology 81:289-304.

Bolanos Gil de Montes F, Perez Resinas FMA, Rodriguez Garcia M, and Gonzalez Ortiz M. 1999. Exophthalmometry in Mexican adults. Rev Invest Clin 51:341-343.

Bothwell RK, Brigham JC, and Deffenbacher KA. 1987. Correlation of witness accuracy and confidence: Optimality hypothesis revisited. J Appl Psychol 72:691-695.

Brennan HG. 1980. Correction of the ptotic brow. Otolaryngol Clin North Am 13:265-273.

Briggs CA. 1998. Anthropological assessment. In Clement JG and Ranson DL (eds.): Craniofacial identification in forensic medicine. London: Arnold, pp. 49-61.

Broca M. 1862. Sur les projections de la tete, et sur un nouvean procede de cephalometrie. Bull. de la Societe d'Anthropologie de Paris 3:514-544.

Bruce V. 1988. Recognising faces. Hove: Lawrence Erlbaum Associates.

Bruce V, Healey P, Burton M, Doyle T, Coombes A, and Linney A. 1991. Recognising facial surfaces. Perception 20:755-769.

Bruce V, and Langton S. 1994. The use of pigmentation and shading information in recognising the sex and identities of faces. Perception 23:803-822.

Bruce V, Valentine T, and Baddeley A. 1987. The basis of the 3/4 view advantage in face recognition. Applied Cognitive Psychology 1:109-120.

Bruce V, and Young A. 1986. Understanding face recognition. B J Psychol 77:305-327.

Bruce V, and Young A. 1998. In the eye of the beholder. New York: Oxford University Press.

Brues AM. 1958. Identification of skeletal remains. Journal of Criminal Law and Criminology and Police Science 48:551-556.

Brues AM. 1992. Forensic diagnosis of race - General race vs specific populations. Soc Sci Med 34:125-128.

Caldwell PC. 1986. New questions (and some answers) on the facial reproduction techniques. In Reichs KJ (ed.): Forensic Osteology. Springfield: Charles C. Thomas, pp. 229-254.

Campsie J. 1997. Maire Claire: Health and Beauty. Sydney: Murdoch Books.

Campsie J. 1998. Maire Claire: Hair and Make-up. Sydney: Murdoch Books.

Cavallaro A, Winzar CF, and Kruger BJ. 1974. The reproducibility of two methods of lateral skull radiography for cephalometric analysis. Aust Dent J 19:122-126.

Chaconas SJ. 1969. A statistical evaluation of nasal growth. Am J Orthod 56:403-414.

Chapple JG, and Stephenson KL. 1970. Photographic misrepresentation. Plast Reconstr Surg 45:135-140.

Cherry DG, and Angel JL. 1977. Personality reconstruction from unidentified remains. FBI Law Enforcement Bulletin 46:12-15.

Clements BS. 1969. Nasal imbalance and the orthodontic patient. Am J Orthod 55:244-264.

Cohn H. 1865. Jahresbericht der Schlesischen Gesellschaft fuer Vaterlandizche Kultur. 43:156.

Cohn H. 1867. Messungen der prominenz der augen mittels eines neuen instruments, des exophthalmometers. Klin Monatsbl f Augenheilk 5:339.

Converse JM. ed. 1977 Reconstructive Plastic Surgery: Principles and Procedures in Correction. Philadelphia: W. B. Saunders.

Cooke MS. 1988. Intersex differences in cranio-cervical morphology and posture in southern Chinese and British Caucasions. Am J Phys Anthropol 77:43-51.

Cooke MS. 1990. Five-year reproducibility of natural head posture: A longitudinal study. Am J Orthod Dentofacial Orthop 97:489-494.

Cooke MS, and Wei SHY. 1988a. An improved method for the assessment of the sagittal skeletal pattern and its correlation to previous methods. Eur J Orthod 10:122-127.

Cooke MS, and Wei SHY. 1988b. The reproducibility of natural head posture: A methodological study. Am J Orthod Dentofacial Orthop 93:280-288.

Craig EA. 2003. Commentry on facial approximation: globe projection guideline falsified by exophthalmometry literature. J Forensic Sci 48:469.

Cutler BL, and Penrod SD. 1995. Mistaken Identification: The eyewitness, pyschology, and the law. New York: Cambridge University Press.

Davanger M. 1970. Principles and sources of error in exophthalmometry. A new exophthalmometer. Acta Opthalmol Scand 48:625-633.

Davies G, Ellis H, and Shepherd J. 1978. Face recognition accuracy as a function of mode of representation. J Appl Psychol 63:180-187.

de Juan E, Hurley DP, and Sapira JD. 1980. Racial differences in normal values of proptosis. Arch Intern Med 140:1230-1.

DiBernardo BE. 1991. Standardized photographs in aesthetic surgery (Letter and Reply). Plast Reconstr Surg 88:373-374.

DiBernardo BE, Adams RL, Krause J, Fiorillo MA, and Gheradini G. 1998. Photographic standards in plastic surgery. Plast Reconstr Surg 102:559-568.

Dickason WL, and Hanna DC. 1976. Pitfalls of comparative photography in plastic and reconstructive surgery. Plast Reconstr Surg 58:166-175.

Diedrich F. 1926. Ein Beitrag zur Prufung der Leistungsfahigkeit der plastischen Rekonstruktionsmethode..'? Deutsche Zeitschrift der Gesellschaft fur Gerichtliche Medizin 8:365-389.

Disaia JP, Ptak JJ, and Achauer BM. 1998. Digital photography for the plastic surgeon. Plast Reconstr Surg 102:569-573.

Dobrostanski T, and Owen CD. 1998. Craniofacial photography in the living. In Clement JG and Ranson DL (eds.): Craniofacial Identification in Forensic Medicine. New York: Oxford University Press, pp. 137-149.

Drews LC. 1957. Exophthalmology. Am J Ophthalmol 43:37-58.

Dumont ER. 1986. Mid-facial tissue depths of white children: An aid in facial feature reconstruction. J Forensic Sci 31:1463-1469.

Dunsky IL. 1992. Normative data for Hertel exophthalmometry in a normal adult black population. Optom Vis Sci 69:562-64.

Edgerton MT, McKnelly LO, and Wolfort FG. 1970. Operating room photography for the plastic surgeon. Plast Reconstr Surg 46:93-95.

Eisenfeld J, Mishelevich DJ, Dann JJ, and Bell WH. 1975. Soft-hard tissue correlations and computer drawings for the frontal view. Angle Orthod 45:267-272.

Ellenbogen R. 1983. Transcoronal eyebrow lift with concomitant upper blepharoplasty. Plast Reconstr Surg 71:490-499.

Ellis HD, Shepherd JW, and Davies GM. 1979. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception 8:431-439.

El-Mehallawi IH, and Soliman EM. 2001. Ultrasonic assessment of facial soft tissue thickness in adult Egyptians. Forensic Sci Int 117:99-107.

Emmert E. 1870. Klin Monastbl f Augenheilk 8:33-42.

Evenhouse R, Rasmussen M, and Sadler L. 1990. Computer aided forensic facial approximation. Biostereometric Technology and Applications 1380:147-156.

Evison MP. 2001. Modeling age, obesity, and ethnicity in a computerized 3-D facial reconstruction. Forensic Science Communications 3: 1-4.

Evison MP, and Green MA. 1999. Presenting three-dimensional forensic facial simulations on the internet using VRML. J Forensic Sci 44:1219-1223.

Farkas LG. 1981. Anthropometry of the Head and Face in Medicine. Oxford: Elsevier.

Farkas LG. 1994a. Examination. In Farkas LG (ed.): Anthropometry of the Head and Face. New York: Raven Press, pp. 3-56.

Farkas LG. 1994b. Photogrammetry of the face. In Farkas LG (ed.): Anthropometry of the Head and Face. New York: Raven Press, pp. 79-88.

Farkas LG. 1994c. Sources of error in anthropometry and anthroposcopy. In Farkas LG (ed.): Anthropometry of the Head and Face. New York: Raven Press, pp. 57-70.

Farkas LG, Bryson W, and Klotz J. 1980. Is photogrammetry of the face reliable? Plast Reconstr Surg 66:346-355.

Farkas LG, Hreczko TM, and Katic M. 1994a. Craniofacial norms in North American Caucasians from birth (one year) to young adulthood. In Farkas LG (ed.): Anthropometry of the Head and Face. New York: Raven Press, pp. 241-336.

Farkas LG, Munro IR, and Kolar JC. 1987. The validity of neoclassical facial proportion canons. In Farkas LG and Munro IR (eds.): Anthropometric Facial Proportions in Medicine. Illinois: Charles C Thomas, pp. 57-66.

Farkas LG, Ngim RCK, and Lee ST. 1994b. Craniofacial norms in 6-, 12-, and 18-year-old Chinese subjects. In Farkas LG (ed.): Anthropometry of the Head and Face. New York: Raven Press, pp. 337-346.

Farrar F. 1977. From skull to visage: a forensics technique for facial restoration. The Police Chief 44:78-80.

Fedosyutkin BA, and Nainys JV. 1993. The relationship of skull morphology to facial features. In Iscan MY and Helmer RP (eds.): Forensic Analysis of the Skull. New York: Wiley-Liss, pp. 199-213.

Ferrario VF, Sforza C, Miani A, and Serrao G. 1995. A three-dimensional evalutation of human facial asymmetry. J Anat 186:103-110.

Fledelius HC, and Stubgaard M. 1986. Changes in eye position during growth and adult life as base on exopthalmometry, interpupillary distance, and orbital distance measurements. Acta Opthalmol Scand 64:481-6.

Fraser I, and Parker D. 1986. Reaction time measures of feature saliency in a perceptual integration task. In Ellis H, Jeeres M, Newcombe F and Young A (eds.): Aspects of Face Processing. Dordrecht: Martinus Nijhoft.

Fricker JP. 1982. Standardized facial photography. Australian Orthodontic Journal 7:168-173.

Fricker JP. 1985. Projected grid facial photography. Aust Dent J 30:414-417.

Galdino GM, Vogel JE, and Vander Kolk CA. 2001. Standardizing digital photography: It's not all in the eye of the beholder. Plast Reconstr Surg 108:1334-1344.

Garlie TN, and Saunders SR. 1999. Midline facial tissue thicknesses of subadults from a longitudinal radiographic study. J Forensic Sci 44:61-67.

Garn SM. 1961. Roentgenographic Cephalometrics. Philadelphia: J.B. Lippincott & Co.

Gatliff BP. 1984. Facial sculpture on the skull for identification. Am J Forensic Med Pathol 5:327-332.

Gatliff BP, and Snow CC. 1979. From skull to visage. J Biocommun 6:27-30.

Gatliff BP, and Taylor KT. 2001. Three-dimensional facial reconstruction on the skull. In Taylor KT (ed.): Forensic Art and Illustration. Boca Raton: CRC Press, pp. 419-475.

Gavan JA, Washburn SL, and Lewis PH. 1952. Photography: An anthropometric tool. Am J Phys Anthropol 10:331-353.

Genecov JS, Sinclair PM, and Dechow PC. 1990. Development of the nose and soft tissue profile. Angle Orthod 60:191-198.

George RM. 1987. The lateral craniographic method of facial reconstruction. J Forensic Sci 32:1305-1330.

George RM. 1993. Anatomical and artistic guidelines for forensic facial reconstruction. In Iscan MY and Helmer RP (eds.): Forensic Analysis of the Skull. New York: Wiley-Liss, pp. 215-227.

George RM, and Singer R. 1993. The lines and grooves of the face: A suggested nomenclature. Plast Reconstr Surg 92:540-542.

Gerasimov M. 1971. The Face Finder. London: Hutchinson & Co.

Gerber FR, Taylor FH, deLevie M, Drash AL, and Kenny FM. 1972. Normal standards for exophthalmometry in children 10 to 14 years of age: Relation to age, height, weight, and sexual maturation. J Pediatr 81:327-329.

Gill GW. 1990. Introduction. In Gill GW and Rhine S (eds.): Skeletal Attributions of Race. Albuquerque: Maxwell Museum of Anthropology, pp. vii-xii.

Gill GW. 1998. Craniofacial criteria in the skeletal attribution of race. In Reichs KJ (ed.): Forensic Osteology: Advances in the identification of human remains. Illinois: Charles C Thomas, pp. 293-317.

Gill GW, and Rhine S. eds. 1990 Skeletal Attributions of Race. Albuquerque: Maxwell Museum of Anthropology.

Gioia VM, Linberg JV, and McCormick SA. 1987. The anatomy of the Lateral Canthal Tendon. Archives of Ophthalmology 105:529-532.

Goldberg RA, Belan A, and Hoenig J. 1999. Relationship of the eye to the bony orbit, with clinical correlations. Australian and New Zealand Journal of Ophthalmology 27:398-403.

Grammer K, and Thornhill R. 1994. Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. J Comp Psychol 108:233-242.

Greyling LM, and Meiring LH. 1992. Morphological study on the convergence of the facial muscles at the angle of the mouth. Acta Anat 143:127-129.

Hackney FL, Nishioka GJ, and Van Sickels JE. 1988. Frontal soft tissue morphology with double V-Y closure following Le Fort I osteotomy. J Oral Maxillofac Surg 46:850-855.

Haglund WD. 1998. Forensic "art" in human identification. In Clement JG and Ranson DL (eds.): Craniofacial Identification in Forensic Medicine. London: Arnold, pp. 235-243.

Haglund WD, and Reay DT. 1991. Use of facial approximation techniques in identification of Green River serial murder victims. Am J Forensic Med Pathol 12:132-142.

Haig ND. 1984. The effect of feature displacement on face recognition. Perception 13:505-512.

Haig ND. 1986. Exploring recognition with interchanged facial features. Perception 15:235-247.

Hancock PJB, Bruce V, and Burton AM. 2000. Recognition of unfamiliar faces. Trends in Cognitive Sciences 4:330-337.

Helmer R. 1984. Schadelidentifizierung durch elekronicshe bildmischung: sugl el beitr zur konstitutionsbiometrie U Dickermessung d Gesichtsweichteile. Heidelberg: Krminalistik-Verlag.

Helmer RP, Rohricht S, Petersen D, and Mohr F. 1993. Assessment of the reliability of facial reconstruction. In Iscan MY and Helmer RP (eds.): Forensic Analysis of the Skull. New York: Wiley-Liss, pp. 229-246.

Henneberg M, Simpson E, and Stephan C. 2002. The human face in biological anthropology: Craniometry, evolution and forensic identification. In Katsikitis M (ed.): The Human Face: Measurement and Meaning. Boston: Kluwer Academic Publishers.

His W. 1895a. Anatomische Forschungen uber Johann Sebastian Bach's Gebeine und Antlitz nebst Bemerkungen uber dessen Bilder. Abh MathPhysikal KI Kgl Sachs Ges Wiss.

His W. 1895b. Johann Sebastian Bach. Forshungen uber dessen Brabstatte, Bebeine und Antlitz. Bericht an den Rath der Stadt Leipzig. Leipzig: FCW Vogel.

Hodson G, Lieberman LS, and Wright P. 1985. In *vivo* measurements of facial tissue thicknesses in American Caucasoid children. J Forensic Sci 30:1100-1112.

Hoffman BE, McConathy DA, Coward M, and Saddler L. 1991. Relationship between the piriform apeture and interalar nasal widths in adult males. J Forensic Sci 36:1152-1161.

Iscan MY. 1993. Introduction of techniques for photographic comparison: Potential and problems. In Iscan MY and Helmer RP (eds.): Forensic Analysis of the Skull. New York: Wiley-Liss, pp. 57-70.

Iscan MY, and Helmer RP. eds. 1993 Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification. New York: Wiley-Liss.

Iscan MY, and Kennedy KAR. 1989. Reconstruction of life from the skeleton. New York: Wiley-Liss.

Jackson DM. 1996. The bone detectives: How Forensic Anthropologists solve crimes and uncover mysteries of the dead. Toronto: Little, Brown & Company.

Jackson E. 1921. The position of the eyeball in the orbit. Trans Sect Ophthalmol AMA 27:58-76.

Janeke JB, and Wright WK. 1971. Studies on the support of the nasal tip. Archives of Otolaryngology 93:458-464.

Kajikawa Y. 1979. Changes in soft tissue profile after surgical correction of skeletal class III malocclusion. Journal of Oral Surgery 37:167-174.

Kesselring UK. 1985. Standards of photography. Plast Reconstr Surg 75:280-281.

Keyser P. 1870. Arch f Augen- un Ohrenheilk 1:183-186.

Keyser P. 1900. Ueber das messen der prominenz des auges. Arch f Augenh 37:1.

Knudtzon K. 1949. On exophthalmometry: The result of 724 measurements with Hertel's exophthalmometer on normal adult individuals. Acta Psychiat Scand 24:523-7.

Kollman J, and Buchly W. 1898. Die Leistungsfahigkeit physiognomischer Rekonstruktionsversuche auf Grundlage des Schadels'. Archiv fur Anthropologie 25:329-359.

Krmpotic-Nemanic J, Kostovic I, Rudan P, and Nemanic G. 1971. Morphological and histological changes responsible for the droop of the nasal tip in advanced age. Acta Otolaryng 71:278-281.

Krogman WM. 1943. Role of the physical anthropologist in the identification of human skeletal remains, Pt. II. FBI Law Enforcement Bulletin 12:12-28.

Krogman WM. 1946. The reconstruction of the living head from the skull. FBI Law Enforcement Bulletin 17:11-17.

Krogman WM. 1962. The Human Skeleton in Forensic Medicine. Illinois: Charles C Thomas.

Krogman WM, and Iscan MY. 1986. The Human Skeleton in Forensic Medicine. Illinois: Charles C Thomas.

Kujawa B, and Strzalko J. 1998. Standard of physical attractiveness. Anthropological Review 61:31-48.

Kustar A. 1999. Facial reconstruction of an artificially distorted skull of the 4th to the 5th century from the site of Mozs. International Journal of Osteoarchaeology 9:325-332.

Langlois JH, and Roggman LA. 1990. Attractive faces are only average. Psychol Sci 1:115-121.

Langlois JH, Roggman LA, and Musselman L. 1994. What is average and what is not average about attractive faces? Psychol Sci 5:214-220.

Lebedinskaya GV, Balueva TS, and Veselovskaya EV. 1993. Principles of facial reconstruction. In Iscan MY and Helmer RP (eds.): Forensic Analysis of the Skull. New York: Wiley-Liss, pp. 183-198.

Lin S-S, and Kerr WJS. 1998. Soft and hard tissue changes in class III patients treated by bimaxillary surgery. Eur J Orthod 20:25-33.

Lindsay DS, Read JD, and Sharma K. 1998. Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. Psychol Sci 9:215-218.

Lindsay RCL, Lea JA, and Fulford JA. 1991a. Sequential lineup presentation: Technique matters. J Appl Psychol 76:741-745.

Lindsay RCL, Lea JA, Nosworthy GJ, Fulford JA, Hector J, LeVan V, and Seabrook C. 1991b. Biased lineups: Sequential presentation reduces the problem. J Appl Psychol 76:796-802.

Lindsay RCL, and Wells GL. 1985. Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. J Appl Psychol 70:556-564.

Little AC, Burt DM, Penton-Voak I, and Perrett DI. 2001. Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. Proc R Soc Lond B Biol Sci 268:1-6.

Macho GA. 1986. An appraisal of plastic reconstruction of the external nose. J Forensic Sci 31:1391-1403.

Macho GA. 1989. Descriptive morphological features of the nose - an assessment of their importance for plastic reconstruction. J Forensic Sci 34:902-911.

Majekodunmi S, and Oluwole M. 1989. Exophthalmometry in Nigerians. West Afr J Med 8:35-37.

Manhein MH, Listi GA, Barsley RE, Musselman R, Barrow NE, and Ubelaker DH. 2000. In vivo facial tissue depth measurements for children and adults. J Forensic Sci 45:48-60.

Marieb RN. 1995. Human Anatomy and Physiology. Redwood City: Benjamin/Cummings Publishing.

McCance AM, Moss JP, Fright WR, Linney AD, and James DR. 1997a. Three-dimensional analysis techniques - Part 1: Three-dimensional soft-tissue analysis of 24 adult cleft palate patients following Le Fort I maxillary advancement: A preliminary report. The Cleft Palate-Craniofacial Journal 34:36-45.

McCance AM, Moss JP, Fright WR, Linney AD, and James DR. 1997b. Three-dimensional analysis techniques - Part 2: Laser scanning: A quantitative three-dimensional soft-tissue analysis using a color-coding system. The Cleft Palate-Craniofacial Journal 34:46-51.

Migliori ME, and Gladstone GJ. 1984. Determination of the normal range of exophthalmometric values for black and white adults. Am J Ophthalmol 988:438-442.

Milner CS, Neave RA, and Wilkinson CM. 2001. Predicting growth in the aging craniofacial skeleton. Forensic Science Communications 3: 1-4.

Montagu A. 1959. Hair growth and hair regeneration: Introduction. Annals of the New York Academy of Sciences 83:362-364.

Montagu MFA. 1947. A study of man embracing error. Technology Review 49:345-347.

Moorees CFA, and Kean MR. 1958. Natural head position: A basic consideration in the interpretation of cephalometric radiographs. Am J Phys Anthropol 16:213-234.

Moorees CFA, van Venrooij ME, Lebret LML, Glatky CB, Kent RLJ, and Reed RB. 1976. New norms for the mesh diagram analysis. Am J Orthod 69:57-71.

Morello DC, Converse JM, and Allen D. 1977. Making uniform photographic records in plastic surgery. Plast Reconstr Surg 59:366-372.

Musch DC, Frueh BR, and Landis JR. 1985. The reliability of Hertel Exophthalmometry. Ophthalmology 92:1177-1180.

Nairn RI. 1976. The concept of occlusal vertical dimension and its importance in clinical practice. In Anderson DJ and Matthews B (eds.): Mastication. Bristol: John Wright and Sons, pp. 58-65.

Nanda RS, Meng H, Kapila S, and Goorhuis J. 1990. Growth changes in the soft tissue facial profile. Angle Orthod 60:177-190.

Neave R. 1998. Age changes to the face in adulthood. In Clement JG and Ranson DL (eds.): Craniofacial Identification in Forensic Medicine. New York: Oxford University Press, pp. 225-231.

Nechala P, Mahoney J, and Farkas LG. 1999. Digital two-dimensional photogrammetry: A comparison of three techniques of obtaining digital photographs. Plast Reconstr Surg 103:1819-1834.

Nelson LA, and Michael SD. 1998. The application of volume deformation to three-dimensional facial reconstruction: A comparison with previous techniques. Forensic Sci Int 94:167-181.

Ngan P, Hagg U, Yiu C, Merwin D, and Wei SHY. 1996. Soft tissue and dentoskeletal profile changes associated with maxillary expansion and protraction headgear treatment. American Journal of Dentofacial Orthopedics 109:38-49.

Nucci P, Brancato R, Bandello F, Alfarano R, and Bianchi S. 1989. Normal exophthalmometric values in children. Am J Ophthalmol 108:582-584.

Oestreicher JH, and Hurwitz JJ. 1990. The position of the eyebrow. Ophthalmic Surgery 21:245-249.

O'Toole AJ, Vetter T, Troje NF, and Bulthoff HH. 1997. Sex classification is better with three-dimensional head structure than with image intensity information. Perception 26:75-84.

Patterson CN. 1980. The aging nose: Characteristics and correction. Otolaryngol Clin North Am 13:275-288.

Peng L. 1999. Fifteen-year reproducibility of natural head posture: a longitudinal study. American Journal of Dentofacial Orthopedics 116:82-85.

Penton-Voak IS, Jones BC, Little AC, Baker S, Tiddeman B, Burt DM, and Perrett DI. 2001. Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. Proceedings of the Royal Society, B 268:1617-1625.

Penton-Voak IS, and Perrett DI. 2000. Female preference for male faces changes cyclically - Further evidence. Evolution and Human Behavior 21:39-48.

Penton-Voak IS, Perrett DI, Castles DL, Kobayashi T, Burt DM, Murray LK, and Minamisawa R. 1999. Menstral cycle alters face preference. Nature 399:741-742.

Perper JA, Patterson GT, and Backner JS. 1988. Face imaging reconstructive morphography: A new method for physiognomic reconstruction. Am J Forensic Med Pathol 9:126-138.

Perrett DI, Lee KJ, Penton-Voak I, Rowland D, Yoshikawa S, Burt DM, Henzill SP, Castles DL, and Akamatsu S. 1998. Effects of sexual dimorphism on facial attractiveness. Nature 394:884-887.

Perrett DI, May KA, and Yoshikawa S. 1994. Facial shape and judgements of female attractiveness. Nature 368:239-242.

Phillips VM, Rosendorff S, and Scholtz HJ. 1996. Identification of a suicide victim by facial reconstruction. J Forensic Sci 14:34-38.

Phillips VM, and Snuts NA. 1996. Facial reconstruction: Utilization of computerized tomography to measure facial tissue thickness in a mixed racial population. Forensic Sci Int 83:51-59.

Popper KR. 1934. Logik der Forschung. Vienna: Julius Springer Verlag.

Popper KR. 1959. The Logic of Scientific Discovery. London: Hutchinson.

Popper KR. 1972. Objective Knowledge: An evolutionary approach. Oxford: Clarendon Press.

Posen J. 1967. Longitudinal study of growth of the nose. Am J Orthod 53:746-756.

Powell N, and Humphreys B. 1984. Proportions of the aesthetic face. New York: Thieme-Stratton.

Prag J, and Neave R. 1997. Making Faces: Using Forensic and Archaeological Evidence. London: British Museum Press.

Prokopec M, and Ubelaker DH. 2002. Reconstructing the shape of the nose according to the skull. Forensic Science Communications 4: 1-4.

Quant JR, and Woo GC. 1992. Normal calues of eye position in the Chinese population of Hong Kong. Optom Vis Sci 69:152-158.

Rabey GP. 1977-78. Current principles of morphanalysis and their implications in oral surgical practice. British Journal of Oral Surgery 15:97-109.

Ras F, Habets LLMH, van Ginkel FC, and Prahl-Andersen B. 1995. Method for quantifying facial asymmetry in three dimensions using stereophotogrammetry. Angle Orthod 65:233-239.

Ras F, Habets LLMH, van Ginkel FC, and Prahl-Andersen B. 1996. Quantification of facial morphology using stereophotogrammetry - demonstration of a new concept. J Dent 24:369-374.

Rathbun TA. 1984. Personal identification: facial reproductions. In Rathbun T and Buikstra J (eds.): Human Identification: Case Studies in Forensic Anthropology. Illinois: Charles C Thomas, pp. 347-362.

Reichs KJ. 1998. Forensic Osteology: Advances in the identification of human remains. Illinois: Charles C Thomas.

Rhine JS. 1990a. Coming to terms with facial reproduction. J Forensic Sci 35:960-963.

Rhine JS, and Campbell HR. 1980. Thickness of facial tissues in American blacks. J Forensic Sci 25:847-858.

Rhine JS, and Moore CE. 1984. Tables of facial tissue thickness of American Caucasoids in forensic anthropology. Maxwell Museum Technical Series 1.

Rhine S. 1990b. Foreword. In Gill GW and Rhine S (eds.): Skeletal Attribution of Race. Albuquerque: Maxwell Museum of Anthropology, pp. xiii-xvii.

Rhodes G, Brennan S, and Carey S. 1987. Identification and ratings of caricatures: Implications for mental representations of faces. Cognitive Psychology 19:473-497.

Rhodes G, Hickford C, and Jeffery L. 2000. Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? B J Psychol 91:125-140.

Rhodes G, Sumich A, and Byatt G. 1999. Are average facial configurations attractive only because of their symmetry. Psychol Sci 10:52-58.

Rhodes G, and Tremewan T. 1996. Averageness, exaggeration, and facial attractiveness. Psychol Sci 7:105-109.

Robinson JM, Rinchuse DJ, and Zullo TG. 1986. Relationship of skeletal pattern and nasal form. Am J Orthod 89:499-506.

Rosenstein TP, Talebzadeh N, and Pogrel A. 2000. Anatomy of the lateral canthal tendon. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 89:24-28.

Rowland DA, and Perrett DI. 1995. Manipulating facial appearance through shape and color. IEEE 15:70-76.

Roy A. 1992. Faceless no more. Reader's Digest:117-122.

Rozprym F. 1934. Eyebrows and eyelashes in man: Their different forms, pigmentation and heredity. The Journal of the Royal Anthropological Institute 64:353-395.

Rubin S. 1998. Murder, he sculpted. Reader's Digest:62-67.

Safe M. 1991. Faceless no more: How an identity was given to a mystery skeleton: The Australian, pp. 14-20.

Sauer NJ. 1992. Forensic anthropology and the concept of race: If races don't exist, why are forensic anthropologists so good at identifying them. Soc Sci Med 34:107-111.

Schlabs G. 1915. Klin Monatsbl f Augenheilk 55:611-635.

Schultz A. 1918. Relation of the external nose to the bony nose and nasal cartilages in Whites and Negroes. Am J Phys Anthropol 1:329-338.

Siersbaek-Nielsen S, and Solow B. 1982. Intra- and interexaminer variability in head posture recorded by dental auxiliaries. Am J Orthod 82:50-57.

Simpson E, and Henneberg M. 2002. Variation in soft-tissue thicknesses on the human face and their relation to craniometric dimensions. Am J Phys Anthropol 118:121-133.

Smith SL, and Buschang PH. 2001. Midsagittal facial tissue thickness of children and adolescents from the Montreal growth study. J Forensic Sci 46:1294-1302.

Snow CC, Gatliff BP, and McWilliams KR. 1970. Reconstruction of facial features from the skull: An evaluation of its usefulness in forensic anthropology. Am J Phys Anthropol 33:221-228.

Solow B, and Tallgren A. 1971. Natural head position in standing subjects. Acta Odontoloiga Scandinavia 29:591-607.

Sporer SL, Read D, Penrod S, and Cutler B. 1995. Choosing, confidence, and accuracy: A meta-anaylsis of the confidence-accuracy relation in eyewitness identification studies. Psychol Bull 118:315-327.

Stadmuller F. 1922. Zur Beurteilung der plastischen Rekonstruktionsmethode der Physiognomie auf dem Schadel. Zeitschrift Morphol. Anthopologie 22:337-372.

Stadmuller F. 1925. Plastische physiognomie-rekonstruktionen auf den beiden diluvialen Schadeln con obercassel bei bonn. Zeitschrift Morphol. Anthopologie 23:301-314.

Steblay N, Dysart J, Fulero S, and Lindsay RCL. 2001. Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. Law and Human Behavior 25:459-473.

Stephan CN, and Henneberg M. 2001. Building faces from dry skulls: Are they recognized above chance rates? J Forensic Sci 46:432-440.

Stephan CN. 2002. Facial Approximation: Falsification of globe projection guideline by exophthalmometry literature. J Forensic Sci 47:1-6.

Stewart TD. ed. 1970 Personal Identification in Mass Disasters. Washingtion: National Museum of Natural History, Smithsonian Institution.

Stewart TD. 1979a. Essentials of forensic anthropology: Especially as developed in the United States. Illinois: Charles C Thomas.

Stewart TD. 1979b. Reconstruction of the facial soft parts. In Stewart TD (ed.): Essentials of Forensic Anthropology. Illinois: Charles C Thomas, pp. 255-274.

Stewart TD. 1983. The points of attachment of the palpebral ligaments: Their use in facial reconstructions on the skull. J Forensic Sci 28:858-863.

Stoney MB, and Koelmeyer TD. 1999. Facial reconstruction: A case report and review of development of techniques. Med Sci Law 39:49-60.

Subtelny JD. 1959. A longitudinal study of soft tissue facial structures and their profile characteristics, defined in relation to underlying skeletal structures. Am J Orthod 45:481-507.

Suk V. 1935. Fallacies of anthropological identifications. Publications de la Facultae des sciences de l'Universitae Masaryk 207:3-18.

Sutton PRN. 1969. Bizygomatic Diameter: The thickness of the soft tissues over the zygions. Am J Phys Anthropol 30:303-310.

Suzuki K. 1948. On the thickness of the soft parts of the Japanese face. J Anthropol Soc Nippon 60:7-11.

Suzuki T. 1973. Reconstitution of a skull. International Criminal Police Review 264:76-80.

Taylor KT. 2001a. Forensic Art and Illustration. Boca Raton: CRC Press.

Taylor KT. 2001b. Two-dimensional facial reconstruction from the skull. In Taylor KT (ed.): Forensic Art and Illustration. Boca Raton: CRC Press, pp. 361-417.

Taylor RG, and Angel C. 1998. Facial reconstruction and approximation. In Clement JG and Ranson DL (eds.): Craniofacial Identification in Forensic Medicine. New York: Oxford University Press, pp. 177-185.

Thomann JV, and Rivett LJ. 1982. Applications of photogrammetry to orthodontics. Australian Orthodontic Journal 7:162-167.

Thomas JR, Tardy ME, and Przekop H. 1980. Uniform photographic documentaiton in facial plastic sugery. Otolaryngol Clin North Am 13:367-381.

Tiddeman B, Burt M, and Perrett D. 2001. Prototyping and transforming facial textures for perception research. IEEE Computer Graphics and Applications 21:42-50.

Tiddeman B, Duffy N, and Rabey G. 2000. Construction and visualisation of three-dimensional facial statistics. Computer Methods and Programs in Biomedicine 63:9-20.

Tiddeman B, Rabey G, and Duffy N. 1999. Synthesis and transformation of three-dimensional facial images: Extending the principles of face-space transormations by using texture-mapped laser-scanned surface data. IEEE Engineering in Medicine and Biology:64-69.

Tiddeman BP. 1998. PhD, Heriot-Watt University, Edinburgh.

Tyrrell J, Evison MP, Chamberlain AT, and Green MA. 1997. Forensic three-dimensional facial reconstruction: Historical review and contemporary developments. J Forensic Sci 42:653-661.

Ubelaker DH. 1993. Facial reconstruction. J Forensic Sci 37:1442-1444.

Ubelaker DH, and O'Donnell G. 1992. Computer-assisted facial reproduction. J Forensic Sci 37:155-162.

van Rensburg J. 1993. Accuracy of recognition of 3-dimensional plastic reconstruction of faces from skulls: Anatomical Society of Southern Africa. Krugersdorp: Game Reserve, pp. 20.

Vanezis P, Blowes RW, Linney AD, Tan AC, Richards R, and Neave R. 1989. Application of 3-D computer graphics for facial reconstruction and comparison with sculpting techniques. Forensic Sci Int 42:69-84.

Vanezis P, Vanezis M, McCombe G, and Niblett T. 2000. Facial reconstruction using 3-D computer graphics. Forensic Sci Int 108:81-95.

Vetter T, Jones MJ, and Poggio T. Year. of Conference A bootstrapping algorithm for learning linear models of object classes. IEEE, Computer Vision and Pattern Recognition, pp. 40-46.

Virchow H. 1912. Die anthropologische Untersuchung der Nase. Zeitschrift fur Ethnologie 44:290-337.

Virchow H. 1924. Zur Anthropologie der Nase. Zeitschrift fur Ethnologie 56:94-121.

von Eggeling H. 1913. Die Leistungsfahigkeit physiognomischer Rekonbstruktionsversuche auf Grundlage des Schadels'. Archiv fur Anthropologie 12:44-47.

Welcker H. 1883. Schiller's Schadel und Todtenmaske, nebst Mittheilungen uber Schadel und Todtenmaske Kant's. Braunschweig: Viehweg F and Son.

White TD. 2000. Human Osteology. San Diego: Academic Press.

Wilkinson CM. 2002. In vivo facial tissue depth measurements for White British children. J Forensic Sci 47:459-465.

Wilkinson CM, and Mautner SA. 2003. Measurement of eyeball protrusion and its application in facial reconstruction [Technical Note]. J Forensic Sci 48:12-16.

Wilkinson CM, Neave RAH, and Smith D. 2002. How important to facial reconstruction are the correct ethnic group tissue depths? Proceedings of the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification, Bari, pp. 111-121.

Wilkinson CM, and Whittaker DK. 2002. Juvenile forensic facial reconstruction - A detailed accuracy study. Proceedings of the 10th Biennial Scientific Meeting of the International Association for Craniofacial Identification, Bari, pp. 98-110.

Williams PL. ed. 1995 Gray's Anatomy, 38th Edition. New York: Churchill Livingstone.

Williamson MA, Nawrocki SP, and Rathbun TA. 2002. Variation in midfacial tissue thickness of African-American children. J Forensic Sci 47:25-31.

Wittbjer J, and Rune B. 1989. Changes of the profile after advancement genioplasty. Scandinavan Journal of Plastic and Reconstructive Surgery 23:65-70.

Yoshino M, and Seta S. 2000. Skull-photo superimposition. In Siegel JA, Saukko PJ and Knupfer GC (eds.): Encyclopedia of Forensic Sciences. San Diego: Academic Press, pp. 807-815.

Zarem HA. 1983a. Standards of photography. Plast Reconstr Surg 74:137-144.

Zarem HA. 1983b. Standards of photography. Plast Reconstr Surg 74:145-146.

Appendix 1: Photography Rig Paper, *Plast Reconstr Surg*

Manuscript in 2nd review

lastic & Reconstructive Surgery
equest for Review

AUG 5 2002

Robert M. Goldwyn, M.D., Editor
1101 Beacon Street, Suite 6E
Brookline, MA 02446
(617)731-8473, fax 731-9580
sjournal@ix.netcom.com

ANUSCRIPT #: 02-ma0669 Date sent: 07/20/2002 Return date: 08/20/2002

EVIEWER:

UTHOR(S): *Carl N. Stephan, John G. Clement, Chris D. Owen, Tad Dobrostanski, Allan Owen*

ITLE: A new rig for standardized craniofacial photography put to the test.

---

rofessional relationships, affiliations or financial interests, which create or pose a conflict of interest for the reviewer, must be disclosed to t ditor. By agreeing to review the manuscript, the reviewer represents that he/she has no relationship or interest, which would conflict with th rforming an unbiased review. If a reviewer has questions about what might constitute a conflict, he/she should contact the Editor.

manuscript submitted for review is a privileged document. As long as it is unpublished, it should remain confidential to the authors, viewers and editors. Do not cite a manuscript or refer to the work or utilize its content before it has been published.

Significance (circle one): **A. Novel findings, new information** [✓ **B. Adds worthwhile substance to old information**] **C. Confirmatory information** **D. Marginal to none**

Similar publications **in PRS (10 years)?** yes **Elsewhere in previous 5 years?** yes

Would the text benefit from shortening? If so, which section(s)?

Style **(circle appropriate categories):** [✓ **A. Clear**] **B. Unclear** **C. Concise** [✓ **D. Verbose**] ++ should reduced 75%

Illustrations:
Are **all the tables/figures essential?** NO **If not, which ones?:** table 1, 2, 3a, 3b, 4a, 4b
A. Indicate figures/tables not suitable for publication.
B. If color, does it add value to the paper?
gs on page 41, 42, 44

References **(please circle):** [✓ ~~**A.**~~ **Adequate**] **B. Inadequate** **C. Missing citations?**
Please supply missing references:

Is **Abstract** adequate? yes

Conclusions (circle one): [✓ **A. Warranted**] **B. Partly justified** **C. More data needed** **D. Not justified—Major flaw**

Is there a question about violation of the Helsinki Doctrine on Human Experimentation?
Circle one: yes [✓ no]

If published, would it warrant a discussion? If so, please name two candidates. Include yourself if appropriate.
1. ______________ 2. ______________

* * *

ORTANCE:
OUTSTANDING
SIGNIFICANT
MODEST
NONE

QUALITY FOR PUBLICATION:
1. ACCEPTABLE AS IS
[✓ 2. REVISION NEEDED]
3. MAJOR REVISION NEEDED
4. REVISION NOT POSSIBLE

* * *

ase provide, on the back of this sheet, useful comments to the author(s) for improving the manuscript. Do not make specific atements about acceptability of the paper. Present criticisms dispassionately since we shall send the author(s) a copy of your aluation.

EASE TYPE YOUR COMMENTS. PLEASE RETURN EVERYTHING TO THE ADDRESS ABOVE AS SOON AS POSSIBLE.
HE EDITORS AND AUTHORS APPRECIATE YOUR HELP IN REVIEWING THIS PAPER.

equest for Review

NOV 12 2002

1101 Beacon Street, Suite 6E
Brookline, MA 02446
(617)731-8473, fax 731-9580
sjournal@ix.netcom.com

ANUSCRIPT #: 02-ma0669

Date sent: 10/20/2002

Return date: ~~08/20/2002~~ 11/10/02

EVIEWER: # 2

UTHOR(S): *Carl N. Stephan, John G. Clement, Chris D. Owen, Tad Dobrostanski, Allan Owen*

TLE: A new rig for standardized craniofacial photography put to the test.

ofessional relationships, affiliations or financial interests, which create or pose a conflict of interest for the reviewer, must be disclosed to th ditor. By agreeing to review the manuscript, the reviewer represents that he/she has no relationship or interest, which would conflict with the erforming an unbiased review. If a reviewer has questions about what might constitute a conflict, he/she should contact the Editor.

manuscript submitted for review is a privileged document. As long as it is unpublished, it should remain confidential to the authors, viewers and editors. Do not cite a manuscript or refer to the work or utilize its content before it has been published.

Significance (circle one): **A. Novel findings, new information** [✓ **B. Adds worthwhile substance to old information**] **C. Confirmatory information** **D. Marginal to none**

Similar publications in PRS (10 years)? yes **Elsewhere in previous 5 years?** yes

Would the text benefit from shortening? YES **If so, which section(s)?** all

Style (circle appropriate categories): [✓ **A. Clear**] **B. Unclear** **C. Concise** **D. Verbose**

Illustrations:
Are all the tables/figures essential? yes **If not, which ones?:**
A. Indicate figures/tables not suitable for publication.
B. If color, does it add value to the paper?

References (please circle): [✓ **A. Adequate**] **B. Inadequate** **C. Missing citations?**
Please supply missing references:

Is **Abstract** adequate? yes

Conclusions (circle one): [✓ **A. Warranted**] **B. Partly justified** **C. More data needed** **D. Not justified—Major flaw**

Is there a question about violation of the Helsinki Doctrine on Human Experimentation?
Circle one: **yes** [✓ **no**]

If published, would it warrant a discussion? If so, please name two candidates. Include yourself if appropriate.
1. ______________________ 2. ______________________

****

PORTANCE :
OUTSTANDING
SIGNIFICANT
MODEST
NONE

QUALITY FOR PUBLICATION:
1. ACCEPTABLE AS IS
[✓ 2. REVISION NEEDED]
3. MAJOR REVISION NEEDED
4. REVISION NOT POSSIBLE

ase provide, on the back of this sheet, useful comments to the author(s) for improving the manuscript. Do not make specific tements about acceptability of the paper. Present criticisms dispassionately since we shall send the author(s) a copy of your aluation.

EASE TYPE YOUR COMMENTS. PLEASE RETURN EVERYTHING TO THE ADDRESS ABOVE AS SOON AS POSSIBLE. E EDITORS AND AUTHORS APPRECIATE YOUR HELP IN REVIEWING THIS PAPER.

REVISED MANUSCRIPT

A new rig for standardized craniofacial photography put to the test

AUTHORS:

Carl N. Stephan, BHSc (Hons)[1,2],

John G. Clement, PhD[2],

Chris D. Owen, BAppSc[2],

Tad Dobrostanski[2],

Allan Owen, DipEng[3]

AUTHOR AFFILIATIONS:

[1] Department of Anatomical Sciences, The University of Adelaide, Australia

[2] School of Dental Science, The University of Melbourne, Australia

[3] School of Engineering, The University of Melbourne, Australia

AUTHORS' CITIES OF WORK:

[1] Adelaide, Australia

[2,3] Melbourne, Australia

RUNNING HEAD:

Standardized Craniofacial Photography

ABSTRACT

This paper describes a photography rig that has been built at the University of Melbourne, Australia, specifically for the purpose of taking rapid and highly standardized craniofacial photographs, in simultaneous views of front and profile. The rig uses a novel projected light range finding system that has been designed for easy and accurate positioning of subjects, in the natural head position, at precise distances from the frontal camera. The results of experiments examining intra-observer error of multiple photographs taken on the rig indicate that high quality, repeatable photographs, can be taken after a reasonably large amount of time has lapsed between photography sessions (5.2 ± 4.2 days). This study also indicates that some variability remains between photographs even when highly standardized protocols are followed. Consequently it is expected that the variation between photographs with limited standardization is much larger and likely to cause significant errors in any comparisons.

INTRODUCTION

Photography is a useful clinical and research tool whose limitations have been extensively discussed in the medical literature. [1, 2] Many specialized photography rigs have also been described. [3-21] However, these papers do not present images or measurements to indicate the repeatability and accuracy of sequential photographs taken using the methods and equipment described. The aim of this paper is to determine the magnitude of intra-observer error that can be expected when taking multiple photographs on a photography rig that is designed for fast, easy, and highly standardized facial photography. The rig enables front and profile views of the face to be captured simultaneously, and uses a novel projected light range finding system that permits subjects to be positioned at constant subject-camera distances in the natural head position.

METHODS

Description of the Photography Rig

The rig is similar to that described by Dobrostanski and Owen [21]. It consists of a rigid aluminum stand that is firmly secured to the floor and supports a frontal and right side profile camera (Fig. 1). The stand is 2138 mm high and has two horizontal beams that are approximately 1650 mm in length. These beams are positioned at 90 degrees in relation to one another (producing an "L" shaped gantry) and each holds a motorized Nikon SLR camera that is vertically mounted. The gantry can be raised or lowered to accommodate subjects as tall as ~2170mm or as short as ~1400mm, however, even shorter subjects can be accommodated with the help of a raised platform. To allow fine adjustment of the height of the cameras, the beams are counterbalanced by a lead weight. An automatic clutch prevents any vertical creep of the beam once the height has been set. Each horizontal beam is fitted with an adjustable horizontal sliding camera platform that enables precise positioning of each camera.

The cameras used

Two Nikon FM-2 35mm bodies, each fitted with an E2 grid-focusing screen and coupled to an MD12 motor drive powered by a 12.8 V, 2 A, DC external power supply, are used. The cameras are fitted with Nikon 105 mm 1:2.8 Macro Nikor lenses. The faces of subjects are positioned in the camera viewfinder at a distance of 1050mm from the camera (1204mm from the film plane) in frontal view, and ~1050mm in profile (~1204mm from the film plane) giving photographs that seem representative of the natural images observed during normal social contact. Subject camera distance is only approximate in profile since the profile camera is fixed and therefore distances vary slightly depending on subject head width. The camera shutters are activated by an infrared triggering system via a handheld, remote, battery-powered, infrared transmitter unit (see inset Fig. 1).

The lighting used

Two Elinchrom Prolinca 2500 self-contained studio electronic flash units are used to illuminate the subject (Fig. 1). To avoid harsh shadows, foil reflectors are used to bounce the light from the flash units towards the subject. One flash unit is positioned 50 degrees to the right of the subject at a distance of approximately 2300mm. The other unit is placed 30 degrees to the left of the subject also at a distance of approximately 2300mm. Both units and reflectors are placed higher than the subject. The slightly closer positioning of the right flash unit results in slight highlights on the subject giving a three-dimensional effect to the final images.[21] This arrangement is sufficient to permit the use of an f-stop of f/16 with 200 ISO Ektachrome E-200 slide reversal film. The use of a focused distance of about 1.2m and an f-stop of f/16 gives a depth of field from about 1.15m to about 1.26m.

Alignment of the subject

A mirror is placed above the lens of the frontal camera and is angled anteriorly and inferiorly at 5 degrees from the vertical so that the subject can see the reflection of their mid-face (including ears) in the mirror and adapt a natural head position, which is reported to be reproducible with little variance.[24, 25] Two incandescent light pointers are used to centrally position the subject within the field of view of each camera. The light pointers are placed on each side of the frontal camera and are angulated horizontally so that the projected beams of light converge at the point of sharpest focus from the cameras. Each light pointer is fitted with 3V Rowi 528 globes powered by an external power supply. The left pointer (from the subject) projects a "<" shaped image, the right a ">" shaped image which form an "x" at the point of sharpest focus. If the two "v" shaped images overlap the subject is too close to the lens and if the two "v" shaped images are spread apart the subject is too far away (Fig. 2). When the camera shutters are activated, the light emitted from the flash units overpowers the pointer lights so the "v's" are not registered on the film. The focus of the range finding pointer lamps can be adjusted to ensure the projected "v's" are crisp at the desired distance.

The profile camera is positioned so that when the subject is correctly aligned, its field of view will include all of the subject's profile (from behind the ear to in front of the nose). To assist positioning of the subject "two cardboard feet" are stuck to the floor to give an approximate indication of the focal planes of the cameras. The wall in front of each camera is painted with a non-glossy medium sky blue paint to give a uniform background that reduces glare and is easily differentiated from the subject in the final images.[4, 16, 19, 21]

Subjects are positioned following methods of Moorees and colleguaes. [13, 21] Subjects are asked to stand on the cardboard feet in an upright position and to look directly at the wall in front of them. At this point the "L" shaped gantry is lowered from its highest position to a position determined by the placement of the light pointers at glabella. The subject is then asked to look at their eyes in the mirror above the lens of the frontal camera. This helps to positions their head in the natural head position if they were not already in it.[21] The subject is then asked to gently "shuffle" forwards or backwards until the pointers form an "x" at the reference point (glabella). Once the subject is correctly positioned the cameras can be activated (simultaneously) from the remote unit without the need for the photographer to view the subject in the camera viewfinder.

In day-to-day use of the rig there is no need to adjust the pre-selected focus of the cameras/pointer aligners because each subject adjusts their distance with respect to the cameras using the range finding light pointers. All that has to be done when photographing multiple subjects is to position the first using the projected light range finding system, activate the shutters and the flash units using the infrared remote, and repeat the process with the next subject. Since the SLR cameras automatically wind on, subject throughput is rapid.

Experiment 1

This experiment was set up to determine the scaling factor required to adjust measurements from the photographs to actual size. A ruler was positioned by fixing it to a camera tripod with Bostik Blu Tack®. The ruler was then photographed and removed by CNS. Upon leaving the room the operator re-entered and repositioned the ruler, photographed it once again, removed it, and left the room. This procedure was repeated until seven sequential photographs were obtained on the same day. Photographs were taken using 200 ISO Ektachrome slide reversal film.

Once the slides were processed they were mounted, and scanned into a computer using a Nikon® SF-2000 slide scanner. The resultant pictures were 1,200 pixels in width, 1,803 pixels in height (48.0 x 72.1mm) and were saved in TIFF format.

The scanned pictures of the ruler were then viewed and measured in Adobe® Photoshop® 6.0. Measurements were made across 100mm sections of the ruler since this distance was relatively large, and is comparable to, or larger than, many measurements of the face such as those of the nose, mouth and eyes. Averaging the measurements, and dividing the true object size (100mm) by this value allowed the calculation the average magnification factor needed to obtain actual size for images placed at the point/plane of sharpest focus.

Experiment 2

The aim of this experiment was to determine what the magnitude of changes in object magnification were, as a result from positioning about the plane of sharpest focus. Since simultaneous profile and frontal images were taken, this value could then be used to adjust for some of the magnification effects when measuring features that fall away from the point of sharpest focus. A ruler (mounted as in experiment 1) was photographed at various positions around the point of sharpest focus. This was done by placing the ruler on a tripod, which had an adjustable upper platform with a stationary base. The platform was moved at 10mm intervals, in various directions, and the ruler photographed. Measurements of the photographed ruler, across 100mm, made it possible to determine changes in magnification due to positioning about the focal plane.

Experiment 3

This experiment was conducted to determine the error in repeated photography of the same human subject on the same day. The subject was positioned as previously described and photographed with a neutral facial expression with his mouth closed, on six occasions. Numerous anatomical landmarks (e.g., ch-ch, cph-cph, al-al, p-p, en-ex, en-en, ex-ex, sa-sa, sba-sba, sa-sba, ls-li, ls-sto, sn-sto, t-ex, t-se, t-g, t-prn, t-pg, g-pg, prn-se, sn-prn, g-sn), as defined by Farkas (1994), were measured on the images and compared to determine the intra-observer error. Error coefficients were generated by dividing the standard deviation by the average and multiplying by one hundred. Superimpositions were also made of the photographs to indicate repeatability.

Experiment 4

This experiment was the same as experiment 3 except that sequential photographs were not taken on the same day (Fig. 3). Photographs were taken, on average, every 5.2 ± 4.2 days.

RESULTS

Measurements of the ruler (across 100mm) indicated that a magnification factor of 49.75124 was needed to achieve actual size in millimeters. Movement of the ruler by 10 mm in any direction resulted in a 1mm change of object magnification (100mm section of the ruler) - equal to a 1% change in object size for every 10mm the object is closer to, or further from, the camera. As expected, the increase was positive when the movement was toward the camera and negative when away from the camera.

Standard deviations for all facial measurements taken on same and different days were small. Coefficients of error were generally less than 3% indicating that the photographs are highly repeatable for both conditions (same and different days). Although the measure for head rotation (the angle formed between t-ex and the horizontal plane in profile) had one of the highest error coefficients (~6.5%) the actual degree of variation for this variable was low (the maximum variation for the angle formed between the horizontal and t-ex was 3.1 degrees, which is comparable to the repeatability of natural head position reported elsewhere [26-29])

Figure 3 shows the six sequential facial photographs taken of the same subject on different days. It can be seen from these images that repeatability is high, although not exact. Photographs taken on the same day showed even higher repeatability as indicated by the "straight" superimpositions displayed in figure 1. Repeatability was generally less for profile images indicating that most variation in subject positioning seemed to occur in the anterior-posterior plane. Superimpositions after alignment on the subject's eyes indicate that the slight deviations in subject positioning appeared to have minimal affects on face morphology (Fig. 3).

DISCUSSION

Overall the results indicate that highly repeatable photographs can be taken on the photography rig using a light range finding system. However, this study demonstrates that even with high standardization, some variations between photographs remain. It is, therefore, expected that between less standardized photographs variability is much greater, especially when subject-camera distances are not precisely controlled by systems like the projected light range finding mechanism described in this paper.

The non-contact properties of the rig are its primary advantages because: (i) subjects are recorded in their natural head position as they normally appear in life; (ii) it allows many subjects to be photographed in minimal time; (iii) the procedure is easy and simple; and (iv) enables photography of smaller children. These items cannot be achieved using contact systems, or other non-contact systems where various instruments are placed about the head.

Another advantage of the rig is that it enables conventional camera bodies to be directly interchanged with digital ones, so could be improved by the installation of digital components. Using conventional SLR's the rig costs approximately US$15, 680 to build but with the installation of say Nikon D1x digital SLR's the cost would approximate US$23,520. Cameras should be specifically ordered in pairs to ensure as much similarity as possible.

While the light range finding systems allows for non-contact subject positioning, care needs to be taken to ensure the "v's" form a precise "x" at the reference point, since small separations/overlaps of the "v's" result in much larger subject displacements. For example, in experiment 2 it was found that when the ruler was displaced posteriorly (or anteriorly) by 10mm the light pointer "v's" were apart (or overlapped) by only 2mm. The light range finding system may be improved by larger separation of the light pointers so that they are more sensitive to subject movement in the antero-posterior plane, however, tradeoffs are needed since if the range finding system is too sensitive subjects may find it difficult to maintain correct position as a result of muscle tone movement.

It also seems that the rig could be further improved by having flash units which adjust their height according to the positioning of the horizontal camera beams so that lighting conditions do not vary between very tall and very short subjects. This would be appear to be significant advantage for longitudinal studies of small children through to adulthood.

SUMMARY

Highly repeatable photographs can be taken on the photography rig described here even when subjects are photographed after a reasonably large amount of time has lapsed between photography sessions (e.g., 30 days). However, this study demonstrates that despite the use of highly standardized methods some variations between photographs remain. It is, therefore, expected that between less standardized photographs variability is much greater, especially when subject-camera distances are not precisely controlled by systems like the projected light range finding mechanism described in this paper.

Corresponding Author: Carl N. Stephan, Department of Anatomical Sciences, The University of Adelaide, Australia, 5005

LITERATURE CITED

1. Gavan, J. A., Washburn, S. L. and Lewis, P. H. Photography: An anthropometric tool. Am J Phys Anthropol 10: 331, 1952.
2. Farkas, L. G. Photogrammetry of the face. In L. G. Farkas (Eds.), *Anthropometry of the Head and Face*, 2nd New York: Raven Press, 1994. Pp. 79-88.
3. Farkas, L. G., Bryson, W. and Klotz, J. Is photogrammetry of the face reliable? Plast Reconstr Surg 66: 346, 1980.
4. Thomas, J. R., Tardy, M. E. and Przekop, H. Uniform photographic documentaiton in facial plastic sugery. Otolaryngol Clin North Am 13: 367, 1980.
5. Morello, D. C., Converse, J. M. and Allen, D. Making uniform photographic records in plastic surgery. Plast Reconstr Surg 59: 366, 1977.
6. Fricker, J. P. Projected grid facial photography. Aust Dent J 30: 414, 1985.
7. Ras, F., Habets, L. L. M. H., van Ginkel, F. C. and Prahl-Andersen, B. Method for quantifying facial asymmetry in three dimensions using stereophotogrammetry. Angle Orthod 65: 233, 1995.
8. Ferrario, V. F., Sforza, C., Miani, A. and Serrao, G. A three-dimensional evalutation of human facial asymmetry. J Anat 186: 103, 1995.
9. Ras, F., Habets, L. L. M. H., van Ginkel, F. C. and Prahl-Andersen, B. Quantification of facial morphology using stereophotogrammetry - demonstration of a new concept. J Dent 24: 369, 1996.

10. Nechala, P., Mahoney, J. and Farkas, L. G. Digital two-dimensional photogrammetry: A comparison of three techniques of obtaining digital photographs. Plast Reconstr Surg 103: 1819, 1999.

11. Thomann, J. V. and Rivett, L. J. Applications of photogrammetry to orthodontics. Australian Orthodontic Journal 7: 162, 1982.

12. Fricker, J. P. Standardized facial photography. Australian Orthodontic Journal 7: 168, 1982.

13. Zarem, H. A. Standards of photography. Plast Reconstr Surg Plast Reconstr Surg 74: 137, 1983.

14. Zarem, H. A. Standards of photography. Plast Reconstr Surg Plast Reconstr Surg 74: 145, 1983.

15. Kesselring, U. K. Standards of photography. Plast Reconstr Surg Plast Reconstr Surg 75: 280, 1985.

16. Galdino, G. M., Vogel, J. E. and Vander Kolk, C. A. Standardizing digital photography: It's not all in the eye of the beholder. Plast Reconstr Surg Plast Reconstr Surg 108: 1334, 2001.

17. Edgerton, M. T., McKnelly, L. O. and Wolfort, F. G. Operating room photography for the plastic surgeon. Plast Reconstr Surg Plast Reconstr Surg 46: 93, 1970.

18. DiBernardo, B. E. Standardized photographs in aesthetic surgery (Letter and Reply). Plast Reconstr Surg Plast Reconstr Surg 88: 373, 1991.

19. DiBernardo, B. E., Adams, R. L., Krause, J., Fiorillo, M. A. and Gheradini, G. Photographic standards in plastic surgery. Plast Reconstr Surg Plast Reconstr Surg 102: 559, 1998.

20. Disaia, J. P., Ptak, J. J. and Achauer, B. M. Digital photography for the plastic surgeon. Plast Reconstr Surg Plast Reconstr Surg 102: 569, 1998.

21. Dobrostanski, T. and Owen, C. D. Craniofacial photography in the living. In J. G. Clement and D. L. Ranson (Eds.), *Craniofacial Identification in Forensic Medicine*, New York: Oxford University Press, 1998. Pp. 137-149.

22. Cooke, M. S. and Wei, S. H. Y. The reproducibility of natural head posture: A methodological study. Am J Orthod Dentofacial Orthop 93: 280, 1988.

23. Moorees, C. F. A., van Venrooij, M. E., Lebret, L. M. L., Glatky, C. B., Kent, R. L. J. and Reed, R. B. New norms for the mesh diagram analysis. Am J Orthod 69: 57, 1976.

24. Moorees, C. F. A. and Kean, M. R. Natural head position: A basic consideration in the interpretation of cephalometric radiographs. Am J Phys Anthropol 16: 213, 1958.

25. Cooke, M. S. Intersex differences in cranio-cervical morphology and posture in southern Chinese and British Caucasions. Am J Phys Anthropol 77: 43, 1988.

26. Peng, L. Fifteen-year reproducibility of natural head posture: a longitudinal study. American Journal of Dentofacial Orthopedics Am J Orthod Dentofacial Orthop 116: 82, 1999.

27. Cooke, M. S. Five-year reproducibility of natural head posture: A longitudinal study. Am J Orthod Dentofacial Orthop 97: 489, 1990.

28. Siersbaek-Nielsen, S. and Solow, B. Intra- and interexaminer variability in head posture recorded by dental auxiliaries. Am J Orthod 82: 50, 1982.

29. Solow, B. and Tallgren, A. Natural head position in standing subjects. Acta Odontoloiga Scandinavia 29: 591, 1971.

FIGURE LEGENDS

FIGURE 1: Photography rig. Inset shows the hand held infrared triggering device. Note that the picture is not to scale. Altered version of image reproduced here with permission from Wiley-Liss (American Journal of Physical Anthropology © 2003).

FIGURE 2: Pointer alignment of subject: (a) too close to camera; (b) correct position; (c) too far from camera. Image reproduced here with permission from Wiley-Liss (American Journal of Physical Anthropology © 2003).

FIGURE 3: Photographs of the same person taken on different days. True image borders are indicated by the black top and right borders while left and bottom borders are artifactual. "Superimposition 1" shows a straight superimposition of the 6 photos taken over different days. "Superimposition 2" shows the superimposition of the 6 different day photos aligned on the eyes. "Superimposition 3" shows a straight superimposition of the sequential photographs taken on the same day. "Superimposition 4" shows the superimposition of same day sequential photographs aligned on the eyes.

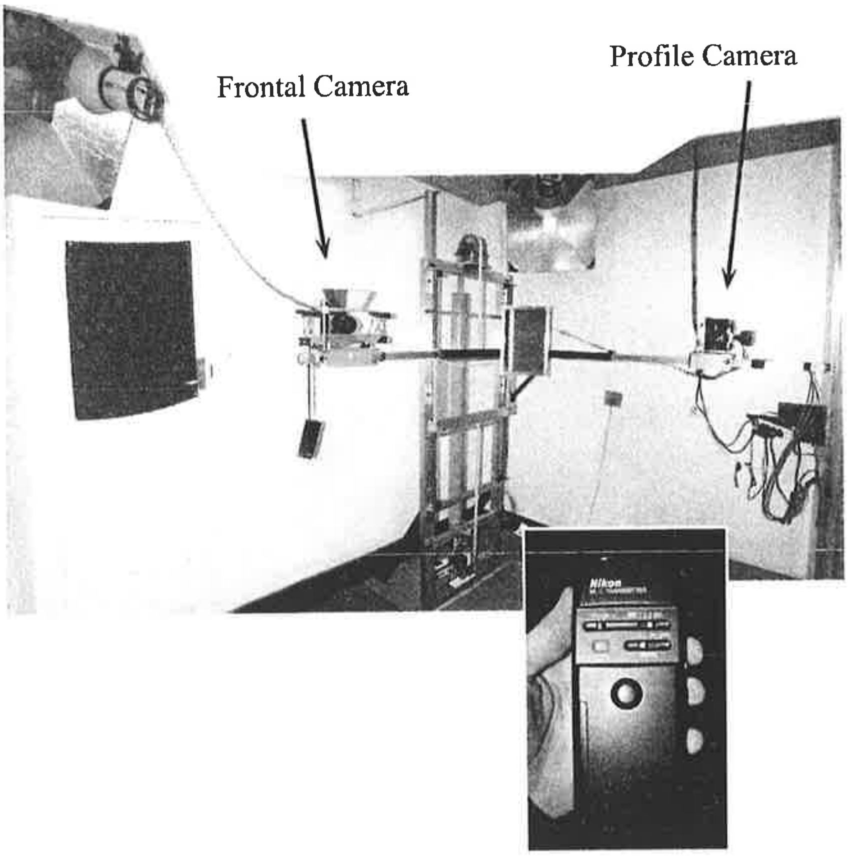
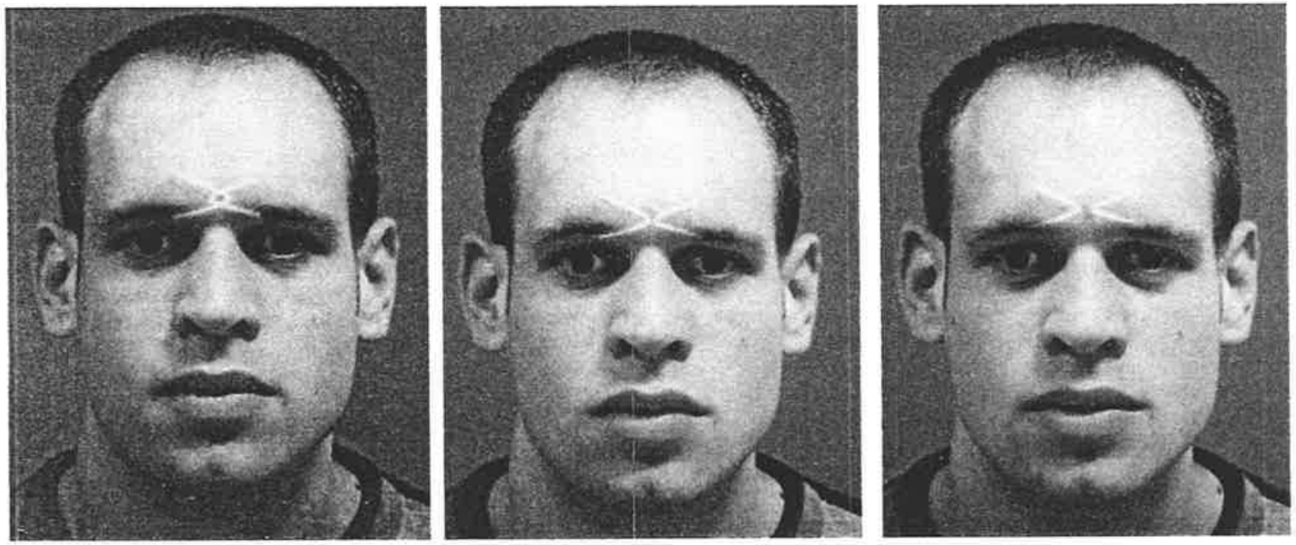


Figure 1



(a) (b) (c)

Figure 2

Day 1

6 days after day 1

7 days after day 1

8 days after day 1

19 days after day 1

30 days after day 1

super-imposition 1

super-imposition 2

super-imposition 3

super-imposition 4

Figure 3

Appendix 2: Mouth Width Paper,
Am J Phys Anthropol

<u>UNCORRECTED</u> Proof

Stephan, C. N. (2003). Facial approximation: an evaluation of mouth-width determination. *American Journal of Physical Anthropology*, *121*(1), 48-57.

Appendix 3: Mouth Width Paper, *J Forensic Sci*

UNCORRECTED Proof

J Forensic Sci, July 2003, Vol. 48, No. 4
Paper ID JFS2002418_484
Available online at: www.astm.org

TECHNICAL NOTE

Carl Stephan,[1] *B.H.Sc. (Hons) and Maciej Henneberg,*[1] *D.Sc.*

Predicting Mouth Width from Inter-Canine Width—A 75% Rule

ABSTRACT: It has been suggested that inter-canine width plus 57% of the cumulative distance between the lateral aspect of the canines and the pupil centers can be used to estimate mouth width (1). Evidence also suggests that the distance between the medial irises approximates actual mouth width fairly well (1). However, these soft tissue prediction guidelines are limited because they rely on accurate medio-lateral positioning of the pupils within the orbits, for which no systematic empirical evidence appears to exist at this stage. It would, therefore, be more appropriate to use only known hard tissue landmarks in mouth width prediction. This study reports the results of using inter-canine width as a percentage of mouth width for its prediction. This method seems favorable in comparison to the other guidelines because it is as accurate, uses known hard tissue landmarks, and does not rely on assumptions concerning pupil location. Estimating mouth width by using the canines alone, therefore, seems the best guideline to use in facial approximation techniques, at least given existing knowledge at this stage.

KEYWORDS: forensic science, facial reconstruction, facial reproduction, soft tissue prediction, hard tissue

Facial approximation is a technique used to build people's faces from their skulls to help establish identification of skeletal remains. Published soft tissue prediction guidelines are often used in the process to help guide practitioners in building the face (2,3); however, many of these guidelines are subjective and have not been systematically evaluated using empirical methods (1).

Traditional techniques used to predict mouth width from the skull include "rules" like mouth width is equal to: (a) pupil width (2,3); (b) medial iris width (4); and (c) width between the lateral aspects of the canines (2,3). Recently, these three soft tissue prediction guidelines have been systematically evaluated using empirical methods (1). This study also proposed an improved guideline that mouth width equals canine width plus 57% of the cumulative distance between the lateral aspect of the canines and the pupil centers on each side of the face (1). The mean error in prediction for this new guideline was found to be 0.1 mm, s.d. 3.4 mm, as compared to a mean error of: 11 mm, s.d. 4 mm for pupil width; 2 mm, s.d. 4 mm for medial iris width; and −13 mm, s.d. 3 mm for inter-canine width (1).

While it seems that the new guideline suggest by Stephan (1) improved upon traditional techniques, the guideline proposed is limited because it relies on pupil positioning that cannot be directly determined from the skull and must consequently be estimated with some unknown error. This limitation also applies to the other reasonably accurate guideline that uses distance between the medial iris borders to determine mouth width. Hence any error in eyeball positioning will result in inaccurate mouth width estimation when using these guidelines (1). Although there seems to be no problem if central positioning of the eyeball is highly accurate, this may not be the case. The central positioning guideline does not seem to have been based on any systematic empirical evidence, but rather on general impressions by Krogman (1962), which may be gross approximations of the truth rather than precise predictions.

Currently no consensus has been reached concerning the accuracy of central eyeball placement in the orbit since few systematic empirical studies have been conducted. Eisenfeld and colleagues (5) attempted to do so but power was low due to small sample sizes ($n = 9$), and hence results must be viewed with caution. Eisenfeld and colleagues (5) found large correlations between inter-pupillary distance and the distance between the centers of the orbits ($r = 0.93$); however, their data suggest a slight overrepresentation of interpupillary distance by central positioning of the pupils within the orbit. Some practitioners, apparently basing their conclusions on past experience, indicate that central positioning is accurate (6–8) or, contrary to Eisenfeld et al. (5), that it results in an underrepresentation of actual interpupillary distance (9). However, these claims should be regarded cautiously since they also appear to be personal impressions that have not been based on formal empirical measurements.

While the accuracy of pupil positioning is unknown, a better approach to predicting mouth width would be to rely on known bony landmarks alone. A simple mouth width prediction guideline that could be used would be calculating mouth width as a percentage of inter-canine width, as opposed to using a direct relationship, which has been shown to be highly inaccurate, i.e., on average results in a 13-mm underestimation (1). Correlations of ~0.50 between mouth width and inter-canine width are reported by Stephan (1), indicating that there may be some value in using the canines as a percentage of mouth width for a prediction guideline.

---

[1] Department of Anatomical Sciences, The University of Adelaide, Australia.

Received 7 Dec. 2002; and in revised form 14 Feb. 2003; accepted 14 Feb. 2003; published XXXX.

It is also reported that inter-canine width is about 75% of actual mouth widths (1); however, error when using this relationship is unknown. This study aims to determine error rates in expressing mouth width as a percentage of inter-canine width in comparison to the more complex guideline (canine width plus 57% of the cumulative distance between the canines and pupils on each side) originally suggested by Stephan (1).

Materials and Methods

Data of Stephan (1) were used here. These data were originally collected using highly standardized photogrammetric methods as described in the original paper (1). Photographs of 93 participants in smiling and relaxed poses were measured for distance between the most lateral aspects of the canines and the width of the mouth (chelion to chelion). The smiling photographs were used for measuring the distance between the most lateral aspects of the canines, and the photographs of participants showing neutral expression were used for measuring mouth width.

Means and standard deviations were calculated for each variable as well as the mean ratio of inter-canine width-to-mouth width. This ratio was used for estimating mouth width, and residuals were calculated along with their standard deviations and compared to the results obtained by Stephan (1) using inter-canine width plus 57% of the cumulative distance between the canines and the pupils. Data were compared using t-tests with statistical significance initially set at the 95% confidence interval but corrected using the Bonferroni adjustment, i.e., $p < 0.05/20 = 0.003$. This adjustment is a conservative one since it takes into account a number of different comparisons across the entire study.

Results and Discussion

The results are presented in Table 1. Overall, inter-canine width averaged 39.5 mm, and mouth width averaged 52.5 mm. Inter-canine width was therefore equivalent to 75.8% of mouth width (or mouth width was about 133% of canine width). Even though canine width (c-c) and mouth width (ch-ch) differed at statistically significant levels between the European sexes (two-tailed, two-sample t-test, $p < 0.001$) and between European and Central/South East Asian females (two-tailed, two-sample t-test, $p < 0.001$), c-c to ch-ch ratios for all samples were fairly consistent. When canine width was used as a percentage to estimate mouth width for the total sample (canine width/0.758), the average residual was −0.2 mm, s.d. 3.5 mm. Table 1 presents data across sub-samples when using the independent canine percentages and when the 0.758 rule is generally determined. It seems that the 75.8% inter-canine width rule worked least best for the "female Central/South East Asian" and the "other individual" groups, suggesting that their independent ratios (79.9 and 73.1%, respectively) may be of some value. However, none of the predicted mouth widths, determined using either the independent or the general percentage guidelines, differed from actual mouth widths at statistically significant levels (two-tailed paired t-tests, $p > 0.003$), indicating that the general guideline is sufficient. While the Bonferroni adjustment used here was rather conservative, we consider the importance of these differences between predicted and actual values to be minimal irrespective of the statistical significance obtained since actual differences in magnitude were small. Further research in this area appears useful since larger sub-samples may act to weaken or strengthen these differences.

These findings indicate that the general inter-canine width percentage guideline (0.758) predicts mouth width essentially as accurately as the other more complex guideline previously suggested by Stephan (1). Average error is barely more when using the inter-canine percentage guideline (−0.2 mm, s.d. 3.5) than when using the more complex rule (1) (0.1 mm, s.d. 3.4 mm). However, the canine width percentage guideline is advantaged because, unlike other guidelines, it does not rely on subjective estimation of pupil location in the orbits. It, therefore, seems more logical to use the distance between the most lateral points of the canines as a percentage since guideline error is similar to that previously obtained and anatomical landmarks used for prediction are known. Since the 95% confidence range of the population mean for the c-c to ch-ch ratio (calculated from the sample mean reported in this study) is from 74.7 to 76.9%, we suggest that it is valid to simply use 75% as the prediction rule, as opposed to 75.8%. This seems useful since 75% is an even number that is easy to remember and apply in practical situations. The adjustment of the ratio by 0.8% slightly increased the inaccuracy of mouth width prediction in the sample reported here, but not by more than 0.6 mm on average for any of the groups studied.

The limitation of using the canines alone, as is the case when using other guidelines, is that asymmetry in horizontal mouth position is not indicated. At this stage, it therefore seems best if the mouth is placed symmetrically over the teeth. As further soft-to-hard tissue relationships are determined, mouth width prediction and positioning accuracy may be increased.

Further studies are required to assess the accuracy of central eyeball placement in the orbit since this rule has not only been used to determine mouth width in some facial approximations but is

TABLE 1—*Means and standard deviations of measurements and calculations made from photographs.*

	Male Central/South East Asian (n=12)		Female Central/South East Asian (n=15)		Male European (n=17)		Female European (n=44)		Other Individuals (n=5)		All Groups (n=93)	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
mouth width (ch-ch)	54.2	5.6	51.2	3.5	55.0	3.2	51.4	3.3	52.5	4.4	52.5	4.0
inter-canine width (c-c)	41.2	1.5	40.8	1.9	41.2	1.5	38.3	1.7	38.4	3.2	39.6	2.2
ratio c-c to ch-ch	76.6	8.2	79.9	5.0	75.2	4.1	74.7	4.4	73.1	2.6	75.8	5.3
mouth width estimation: c-c as a % of ch-ch (=c%)	53.7	1.9	51.0	2.4	54.9	2.0	51.3	2.3	52.4	4.4	52.3	2.9
Average residual of c% to ch-ch	−0.5	5.6	−0.2	3.1	−0.1	2.9	−0.1	2.9	0.0	1.7	−0.2	3.5
mouth width estimation: c-c/0.758	54.3	1.9	53.8	2.5	54.4	1.9	50.5	2.2	50.6	4.2		
Average residual of c-c/0.758 to ch-ch	0.1	5.6	2.6	3.1	−0.6	2.9	−0.9	2.9	−1.9	1.7		

practically used by almost all practitioners for its primary purpose—positioning the eyes. The finding that traditional eyeball positioning in the antero-posterior plane is inaccurate (9), and even misreported (10,11), suggests that other eyeball positioning guidelines may also need to be reevaluated and reassessed.

As indicated by several authors many years ago (12–14), there is a clear need for much of the facial approximation technique to be systematically and empirically evaluated. It seems that a general lack of methodological rigor in the past has lead to the use of many guidelines that have not been formally tested apart from soft tissue depths. Consequently, the claimed accuracy of many soft-tissue prediction guidelines remains to be demonstrated (as does the overall accuracy of the method). The lack of critical reviews and empirical tests of facial approximation methods may have even led to misleading quotes of facial approximation guidelines in the literature (10,11). To date, the results of empirical studies specifically testing facial approximation methods (1,9,15,16) have justified concerns that some subjective guidelines used are not very accurate or reliable and have provided support to critical arguments. However, some progress seems to be being made in testing and developing new guidelines for which error rates are known. Ideally this may lead to facial approximation techniques that are more accurate and reliable in the future.

References

1. Stephan CN. Facial approximation: an evaluation of mouth width determination. Am J Phys Anthropol. In press.
2. Krogman WM. The human skeleton in forensic medicine. Illinois: Charles C. Thomas, 1962.
3. Krogman WM, Iscan MY. The human skeleton in forensic medicine. Illinois: Charles C. Thomas, 1986.
4. Prag J, Neave R. Making faces: using forensic and archaeological evidence. London: British Museum Press, 1997.
5. Eisenfeld J, Mishelevich DJ, Dann JJ, Bell WH. Soft-hard tissue correlations and computer drawings for the frontal view. Angle Orthod 1975;45:267–72.
6. Gatliff BP. Facial sculpture on the skull for identification. Am J Forensic Med Pathol 1984;5:327–32.
7. Gatliff BP, Taylor KT. Three-dimensional facial reconstruction on the skull. In: Taylor KT, editor. Forensic art and illustration. Boca Raton: CRC Press, 2001:419–75.
8. Taylor KT. Forensic art and illustration. Boca Raton: CRC Press, 2001.
9. Stephan CN. Facial approximation: falsification of globe projection guideline by exophthalmometry literature. J Forensic Sci 2002;47:1–6.
10. Craig E. Commentary on facial approximation: globe projection guideline falsified by exophthalmometry literature. J Forensic Sci. In press.
11. Stephan C. Commentry on facial approximation: globe projection guideline falsified by exophthalmometry literature. J Forensic Sci. In press.
12. Suk V. Fallacies of anthropological identifications. Publications de la Facultae des sciences de l'Universitae Masaryk 1935;207:3–18.
13. Montagu MFA. A study of man embracing error. Technol Rev 1947; 49:345–7.
14. von Eggeling H. Die Leistungsfahigkeit physiognomischer Rekonbstruktionsversuche auf Grundlage des Schadels'. Archiv fur Anthropologie 1913;12:44–7.
15. Farkas LG, Munro IR, Kolar JC. The validity of neoclassical facial proportion canons. In: Farkas LG, and Munro IR, editors. Anthropometric Facial Proportions in Medicine. Illinois: Charles C. Thomas, 1987: 57–66.
16. Stephan CN. Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline. Forensic Sci Int 2002; 130:29–33.

Additional information and reprint requests:
Carl Stephan
Department of Anatomical Sciences
University of Adelaide, Adelaide 5005
Australia
E-mail: carl.stephan@adeliade.edu.au

Appendix 4: Pronasale Position and Nose Projection Paper, *Am J Phys Anthropol*

UNCORRECTED Proof

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 121:000–000 (2003)

Predicting Nose Projection and Pronasale Position in Facial Approximation: A Test of Published Methods and Proposal of New Guidelines

Carl N. Stephan,[1]* Maciej Henneberg,[1] and Wayne Sampson[2]

[1]*Department of Anatomical Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia*
[2]*Dental School, University of Adelaide, Adelaide, South Australia 5005, Australia*

KEY WORDS forensic science; facial reconstruction; facial reproduction; soft-tissue prediction guideline; hard tissue; estimation

ABSTRACT Many prediction guidelines exist in facial approximation for determining the soft-tissue features of the face, and the reliability of each is generally unknown. This study examines four published and commonly used soft-tissue prediction guidelines for estimating nose projection, two of which also estimate the position of the pronasale. The methods tested are those described by: 1) Gerasimov ([1971] The Face Finder; London: Hutchinson & Co.), using the distal third of the nasal bones and the nasal spine; 2) Krogman ([1962] The Human Skeleton in Forensic Medicine; Springfield: Charles C. Thomas), using the average soft-tissue depth at midphiltrum, plus three times the length of the nasal spine (and a variation of this technique: plus three times the distance of the tip of the nasal spine from the nasal aperture); 3) Prokopec and Ubelaker ([2002] Forensic Sci Commun *4:*), using the reflected profile line of the nasal aperture; and 4) George ([1987] J Forensic Sci *32:*1305–1330), using a variation of the Goode method. Four identical hard-tissue tracings were made of 59 adult lateral head cephlograms (29 males, mean age 24, SD 10 years; 30 females, mean age 23, SD 5 years) on separate sheets of tracing paper. One soft-tissue tracing was also made for each radiograph. All tracings were marked with three identical reference points. Soft-tissue tracings were isolated from one of us (C.N.S.), who attempted under blind conditions to predict pronasale position and nose projection on the hard-tissue tracings, using the soft-tissue prediction guides above. Actual soft-tissue tracings were then compared to each of the predicted tracings, and differences in projection/pronasale position were measured. Results indicate that for nose projection, methods 3 and 4 performed well, while methods 1 and 2 performed poorly. Features which are most related to nose projection/pronasale are described in this paper, as are regression equations generated from these variables that predict pronasale/nose projection better than the traditional methods mentioned above. The results of this study are significant because they: 1) indicate that the popular facial approximation methods used to build the nose are inaccurate and produce incorrect nose anatomy; and 2) indicate that the new pronasale prediction methods developed here appear to have less error than traditional methods. Am J Phys Anthropol 121: 000–000, 2003. © 2003 Wiley-Liss, Inc.

Facial approximation is the process of estimating the soft tissues of the face from the skull (Prag and Neave, 1997; Taylor, 2001). The aim of forensic facial approximation is to generate specific and purposeful recognition of the target individual (individual to whom the skull belongs) from the advertisement of images of the constructed face in the media (Prag and Neave, 1997; Taylor, 2001). Facial approximation has successfully generated tentative identifications in some forensic cases (Cherry and Angel, 1977; Farrar, 1977; Gatliff and Snow, 1979; Perper et al., 1988; Phillips et al., 1996; Prag and Neave, 1997; Rathbun, 1984; Stoney and Koelmeyer, 1999; Suzuki, 1973).

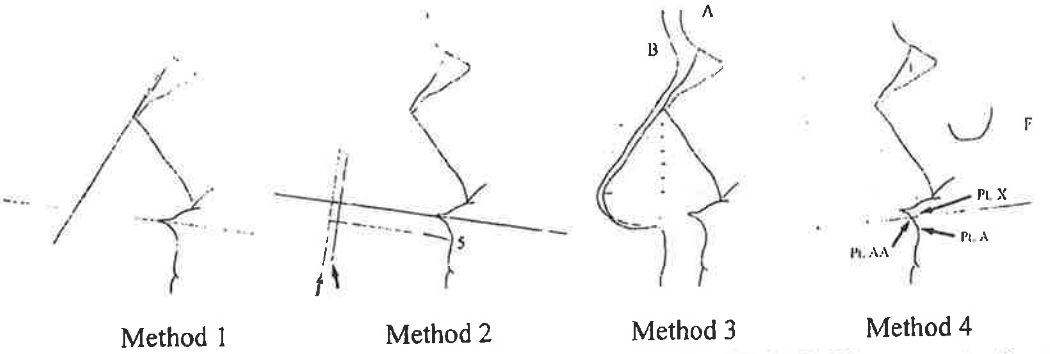
Although facial approximation appears successful in some cases, the ability for facial approximation to achieve its goal is contentious (Brues, 1958; Haglund and Reay, 1991; Montagu, 1947; Stephan, 2002a–c; Stephan and Henneberg, 2001; Suk, 1935). Those who believe the technique works cite instances where facial approximation has been successful (e.g., Prag and Neave, 1997) and make comparisons between the facial approximation and the target individual, pointing out similarities (e.g., Helmer et al., 1993; Krogman, 1946; Prag and Neave, 1997; Suzuki, 1973; Taylor, 2001).

---

*Correspondence to: Carl N. Stephan, Department of Anatomical Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia. E-mail: carl.stephan@adelaide.edu.au

Received 13 August 2002; accepted 21 February 2003.

DOI 10.1002/ajpa.10300



Method 1 Method 2 Method 3 Method 4

Fig. 1. Pronasale/nose-projection prediction according to four methods, using male adult skull of European extraction as example. For actual soft-tissue to hard-tissue relationship, see Figure 4. Method 1: Gerasimov (1971). Two tangents are projected, one following last (distal) third of nasal bones, and the other following general direction of nasal spine. Where tangents intersect indicates tip of nose (pronasale). When placing nasal bone tangent, we gave priority to superior surface of nasal bones, as would be done with real skulls where complete nasal bone profiles cannot usually be considered due to visualization limitations. Method 2: Krogman (1962). Three times length of nasal spine (measured from junction of vomer and maxilla to tip of nasal spine; Taylor, 2001) is added to average soft-tissue depth (according to Rhine and Campbell, 1980; Rhine and Moore, 1984), as suggested by Taylor, 2001) at midphiltrum (defined as "midline point placed as high as possible before curvature of anterior nasal spine begins" (Taylor, 2001, p. 354), following general direction of nasal spine (Taylor, 2001). At point of predicted nose projection, a perpendicular was drawn to indicate level of nose projection (black arrow), as indicated by Taylor (2001). Also tested was a variation of this guideline by substituting original nasal spine length for that determined by margin of most prominent lateral nasal aperture line to tip of nasal spine (nose estimation using this guideline indicated by gray arrow; see Materials and Methods for further explanation). Method 3: Prokopec and Ubelaker (2002). Nasal aperture is divided into seven equal segments along a line at rhinion (line B) that is parallel to nasion-prosthion plane (line A). Distance from line B to edge of aperture is measured and mirrored anteriorly. Two millimeters are added to each of these measures; midline average soft-tissue depths at nasion, midnasal bone, rhinion, and subnasale (according to Helmer, 1984) are added, and points are joined, allowing profile shape of nose, and hence its projection, to be estimated (Prokopec and Ubelaker, 2002). Method 4: George (1987). Distance from nasion to point A (point of most flexion on maxilla in profile) is measured. A percentage of this distance (60.5% for males and 56% for females) is represented along a line parallel to Frankfurt horizontal (line F) (R. George, 2002, personal communication) beginning at point (Pt.) X, which falls on nasion-point A plane at height of point AA (a point located midway along inferior slope of anterior nasal spine). A line perpendicular to Frankfurt horizontal is placed at level of predicted nose projection.

However, the similarity between facial approximations and target individuals has been shown not to indicate the accuracy of facial approximations (ability to be recognized as a target individual) (Stephan, 2002a). Casework success also seems unrepresentative of actual facial approximation success, since many instances of nonsuccess may go unreported (Stephan and Henneberg, 2001), and casework success may be due to factors independent of the facial approximation, such as contextual information (Haglund and Reay, 1991) or chance (Stephan and Henneberg, 2001). Furthermore, recent systematic evaluations of the method have resulted in infrequent true-positive recognitions of the target individual (Stephan and Henneberg, 2001), and when positive identifications have occurred the recognition rates have been low, usually much less than 54% above chance (Snow et al., 1970; Stephan and Henneberg, 2001; van Rensburg, 1993). The rare true-positive recognition of facial approximations and their low recognition rates when they are identified may be due to the use of inaccurate soft-tissue prediction guidelines. This is supported by recent evidence that found specific soft-tissue prediction methods to be inaccurate (Farkas et al., 1987; Stephan, 2002b–d). Additionally, the fact that many *disparate* methods are commonly reported for predicting the *same* facial feature (e.g., the four methods for describing nose projection; see below) suggests that neither the soft to hard tissue relationships, nor the reliabilities of at least some of these techniques, are known (Stephan and Henneberg, 2001). Otherwise, would not the most accurate technique be disproportionately cited and used? It therefore seems likely that some of these other untested soft-tissue prediction guidelines are also inaccurate. This study examines several soft-tissue prediction guidelines used to estimate nose projection and pronasale position.

Although the nose has been regarded as being of little import for face recognition in frontal view (Fraser and Parker, 1986; Haig, 1984, 1986), its contribution to recognition may increase in other views, e.g., profile (Bruce, 1988). As a result, the nose appears to be an important character to predict accurately in facial approximation.

There are four published methods commonly being used to predict nose projection, most of which appear to have been subjectively determined (Fig. 1). It appears that the authors first proposing these F1 techniques (or at least first publishing them) were: Gerasimov (1971) for methods 1 and 3, Krogman (1962) for method 2, and George (1987) for method 4. Gerasimov (1971) originally proposed method 1 for determining pronasale position and nose projection, while he used method 3 to determine the profile shape of the nose. However, following the descriptions by Prokopec and Ubelaker (2002), it seems method 3 can also be used to indicate nose projection (Fig. 1). Since Prokopec and Ubelaker (2002) provided a detailed description of method 3, in this study they will be linked with the technique, even though Gerasimov (1971) originally developed it. Methods 1 and 2 seem the most popular of the four methods, having been used by authors in many leading texts, e.g., method 1: Fedosyutkin and Nainys (1993); Gerasimov (1971); and Prag and Neave (1997); and method 2: Gatliff (1984); Gatliff and Taylor (2001); and Taylor (2001). Despite the publication and broad use of all four methods of nose prediction, no accuracy/reliability tests have been published, and it therefore appears that at least some practitioners have blindly followed the informal observations of others. Although this is unavoidable in many instances, it is not favorable, since these soft-tissue prediction guidelines may be far from the truth.

Virchow (1912, 1924) apparently "emphasized that the prominence of the nose cannot a priori be predicted by the bony nose" (Macho, 1986, p. 1392). However, Gerasimov (1971, p. 54) reported that method 1 was established as being reliable after "many years of work," and that this guideline will generally give the position of the tip of the nose, but no evidence is cited. Prag and Neave (1997), in contrast to Gerasimov (1971), suggested that method 1 only approximates nose projection rather than determines it reliably, but again no evidence is provided. Stephan and Henneberg (2001) provided several examples in support for their statement that method 1 results in an overprediction of nose projection, and unrealistically large noses. It seems likely that method 4 may not be accurate in predicting actual nose projection (R. George, 2002, personal communication) because the method is based on the "aesthetic" surgical methods of Goode that describe a "balanced" nasal projection when the nasofacial angle is at about 36° (Powell and Humphreys, 1984), which many individuals are not expected to possess (note: "balanced" and "aesthetic" are subjectively loaded terms). However, in the three case examples presented by George (1987), method 4 appeared to work fairly well. Taylor (2001) indicated that method 3 does not have any sound basis in anatomy, yet used method 2, which may also be questioned on similar grounds, since no reliable and/or specific hard/soft-tissue relationships, at the present time, are known to exist between the length of the nasal spine and the projection of the soft tissue of the nose.

Some authors reported significant correlations between nose projection and the angle of the nasal bones (e.g., r = 0.35 [Chaconas, 1969]), suggesting that the angle of the nasal bones may be useful in determining nose projection/pronasale position. However, while many (e.g., Fedosyutkin and Nainys, 1993; Gerasimov, 1971; Prag and Neave, 1997; Taylor, 2001) suggest that the nasal spine indicates if the nose points up or down or is horizontal and use this relationship to predict nose projection/pronasale position, no formal tests appear to have been published. While it may be correct in a very general sense, using a direct relationship as in method 1 appears to be, at least, ambitious, given that most nose tips are located considerably above the nasal spine, as easily seen from lateral head radiographs.

The aim of this study is to test, using lateral head cephalograms, the accuracy of the four methods traditionally used in facial approximation to predict nose projection and/or pronasale position (as described in Fig. 1). By doing so, we intend to quantify error in these guidelines, which in the context of findings from other tests of soft-tissue prediction guidelines should give a better indication of the overall accuracy of the facial approximation method. Additionally, we also intend to improve nose projection guidelines if possible, and to help increase the accuracy of facial approximation methods in general.

MATERIALS AND METHODS

The sample consists of 29 male (mean age, 24 years; SD, 10 years) and 30 female (mean age, 23 years; SD, 5 years) lateral head cephalograms of Australians of European extraction, randomly selected from the records of the Adelaide Dental Hospital and the University of Adelaide. While this sample may be biased somewhat toward patients thought to require orthodontic treatment, the sample also includes individuals who were not in need of treatment and who had been x-rayed as part of a checkup. Since all individuals were drawn from the general population and may be subject to a forensic enquiry, no attempt was made to limit the sample to what may be considered "normal." Thus, no individuals were excluded if they displayed malocclusions, prognathism/retrognathism, or missing teeth, and/or were pre/postoperative.

Five tracings were made for each lateral cephalogram. First, one soft-tissue and one hard-tissue tracing were made. Then, three duplications of the hard-tissue tracing were made, giving four identical dependent hard-tissue tracings and one soft-tissue tracing in total for each lateral cephalogram. Tracings were made using a 0.5-mm HB retractable pencil, and a fluorescent light box in a darkened room with aperture style shields to remove excess light generated by the light box. Each individual's tracings were marked with three identical reference points, so that precise superimpositions could be made at a later time. The soft-tissue tracing was isolated from the other four tracings, which C.N.S. used to estimate nose projection, under blind conditions, according to methods described in Figure 1.

Tracing and measuring accuracies were assessed by comparing measurements of tracings and independent retracings of 10 individuals. We chose to examine measurements that would be used in the main experiment. The technical error of tracing/measurement was assessed as a coefficient of variation of the error (CVE). The CVE was calculated by taking the sum of the squared differences between

TABLE 1. Technical error of tracing/measurement assessed as a coefficient of variation of error (CVE)

	CVE
x	0.01
y	0.01
Nasion to rhinion	0.04
Nasion to point A	0.02
Nasion to point AA	0.01
Anterior tip of nasal spine to profile line of nasal aperture	0.09

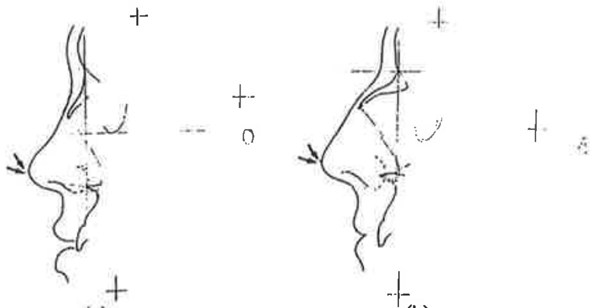

(a) (b)

Fig. 2. Selected superimpositions of independent tracings made from same radiograph. Note placement of pronasale in each case, indicated by dash and highlighted by arrows. **a:** High repeatability superimpositions, although there is some difference in inferior aspect of nasal bone and inferior aspect of nasal spine curve. **b:** Low repeatability superimpositions, with differences between representation of external auditory meatus, orbit margin, inferior aspect of nasal spine curve, and midportion of nasal aperture profile.

test and retest and dividing it by 2× the number of remeasured individuals. The square root of the result was taken and divided by the mean of the test/retest result of the first individual. The CVE were low for measurements of the tracings, indicating that the combined tracing and measuring techniques were rather repeatable (Table 1). Areas of greatest variation appeared to be in determining/tracing the orbit, external auditory meatus, lower anterior profile line of the nasal aperture, and shape of the inferior aspect of the nasal spine. Figure 2 shows superimpositions of two trace/retraces that indicate the high and low extremes of repeatability observed. Over the 10 independent retraces, Frankfurt horizontal repeatability was fairly high, with the difference between tracings being on average 0°, with a standard deviation of 1°.

A variation of method 2 (plus three times the distance from the tip of the nasal spine to the border of the nasal aperture at its base) was included, since on lateral cephalograms the junction of the vomer with the maxilla can be indistinct (Fig. 3), and it was thought that the aperture borders probably fell close to the vomer/maxilla junction. However, we have no direct evidence for this relationship. Although the outline of the vomer bone could be determined on some radiographs, it was generally difficult, and the precise point of the vomer/maxilla junction was also difficult to establish, since few shape or density differences were evident between the two bones in profile. Figure 3 shows the delineation of the nasal septum, and the relative position of the vomer/maxilla junction on an x-ray of a dry skull. The anterior aspect of the perpendicular ethmoid plate was generally difficult to determine in the lateral head radiographs due to hard- and soft-tissue shielding, as was the upper portion of the vomer bone. However, the distal portion of the vomer bone was visible on 44% of cephalograms, and when it was, the length of the nasal spine to the vomer/maxilla junction was estimated. This gave a sample size of 14 females (mean age, 23 years; SD, 5 years) and 12 males (mean age, 28 years; SD, 15 years; range, 19–72 years). The border of the nasal aperture was also used for measuring nasal spine length in these and all other cephalograms.

Once the soft-tissue nose projection had been estimated on the hard-tissue tracings, the actual soft-tissue tracings were superimposed. Measurements were made for actual and predicted nose projection/pronasale position, using a Cartesian axis set up about the Frankfurt horizontal (x) and a perpendicular at the nasion (y) (Fig. 4). The most anterior point of the nose (i.e., pronasale/nose projection) was determined by bisecting the curve of the anterior-most tip of the nose, as determined by sliding a perpendicular anteriorly along the plane of the Frankfurt horizontal. Additional features to those used in estimating nose projection/pronasale were also measured to determine if they explained any variance in pronasale position. These included: the length of the nasal bones from the nasion to rhinion; and the angle of the nasal bones from the nasion to rhinion with respect to the Frankfurt horizontal (Fig. 4). The angle of the nasal spine and that of the soft-tissue nasal septum were also measured with respect to the Frankfurt horizontal (Fig. 4). All distance measures, but not angles, were reduced by a factor of 0.088 to obtain actual values, since the cephalogram images were magnified by this factor in comparison to life.

F4

The predicted and actual measures of nose projection/pronasale position were compared statistically using paired and unpaired two-tailed t-tests with significance levels set at $P < 0.05$, but altered according to the Bonferroni adjustment where necessary. Other measures were analyzed using scatterplots and regression, to try to determine more accurate ways of predicting nose projection and pronasale position. All data analyses were conducted using Microsoft Excel®.

RESULTS

Actual pronasale position averaged, in males, x = 30.9 mm, SD 4.7 mm; y = 44.3 mm, SD 3.5 mm; and in females, x = 28.1 mm, SD 4.1 mm; y = 43.0 mm, SD 3.5 mm. Nose projection (x) was different at statistically significant levels between the sexes ($P < 0.05$), while pronasale height (y) did not differ statistically ($P > 0.05$).

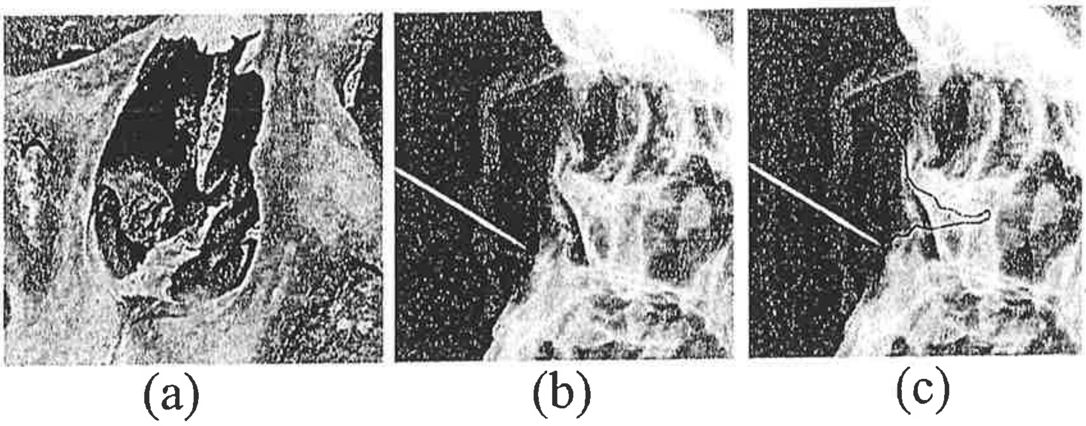


Fig. 3. Nasal aperture and nasal septum delineation in lateral x-ray of dry skull. **a:** Oblique photographic image of nasal aperture, vomer bone, and perpendicular ethomoid plate. **b:** Lateral x-ray of skull, with pin indicating vomer/maxilla junction held in place by Bostik Blu Tak®. **c:** Anterior profile line of nasal septum (vomer bone and perpendicular ethmoid plate).

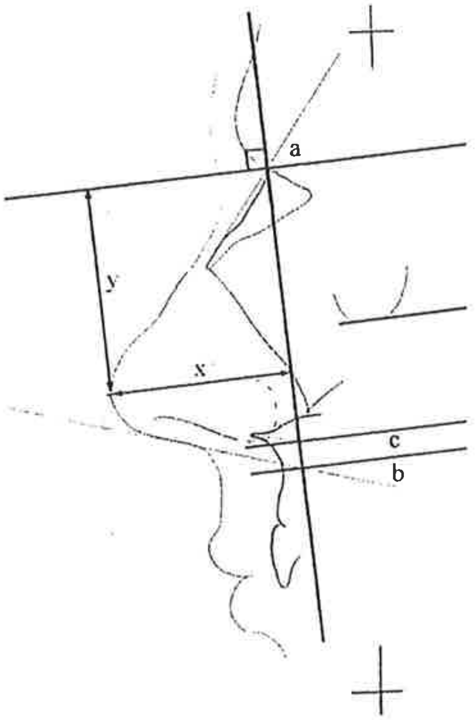


Fig. 4. Measurement of pronasale/nose projection and additional angles. Black lines indicate those in same plane as Frankfurt horizontal or perpendicular to it. x, pronasale projection; y, pronasale "height;" a, nasal bone angle; b, soft-tissue nasal septum angle; c, nasal spine angle.

Table 2a summarizes the accuracy of methods 1, the variation of 2, 3, and 4. Table 2b summarizes the accuracies of method 2, using the vomer/maxilla junction instead of the profile line of the nasal aperture. Overall, methods 4 (George) and 3 (Prokopec/Ubelaker) estimated nose projection best (error of about 2 mm, SD 4 mm), with method 4 having slightly lower standard deviations of error than method 3 (Table 2a,b). The other methods performed much worse, having either large average errors and/or large standard deviations of error. Method 1 (Gerasimov) performed worst of all, having an average error of ~5 mm and a standard deviation of error equal to ~9 mm (Table 2a,b). Overall, the methods predicted nose projection slightly better in males, as estimated values were less often statistically different from actual values, and had lower mean errors than females (Table 2a,b).

Table 3 summarizes the accuracies of methods 1 and 3, used to estimate pronasale position. Accuracy for method 3 (Prokopec/Ubelaker) was modest, having an error (shortest distance between the predicted position of pronasale and the actual position) of about 5 mm, with a standard deviation of about 2 mm (Table 3). Method 1 (Gerasimov) performed poorly, having an average error of about 11 mm, with a standard deviation of about 8 mm. Overall, most predicted averages were found to differ at statistically significant levels from actual averages (Tables 2 and 3).

Nasal bone angle, measured between a line from the nasion to rhinion and the Frankfurt horizontal, was not related to soft-tissue nose height measured vertically from the pronasale to the level of nasion (males $r^2 = 0.00$, females $r^2 = 0.08$). But nasal bone angle was related to nose projection (males $r^2 = 0.52$, females $r^2 = 0.54$). These relationships were slightly stronger than those between nose projection and the angle formed by the distal 1/3 of the nasal bones (males $r^2 = 0.32$, females $r^2 = 0.51$). The distance between the tip of the nasal spine and the border of the nasal aperture at its base also explained some of the variance in nose projection in both males ($r^2 = 0.35$) and females ($r^2 = 0.15$). For males, the horizontal distance from the rhinion to the most posterior point on the nasal aperture profile was related to nose projection ($r^2 = 0.20$). Using these three variables for males ($r^2 = 0.66$) and two for females ($r^2 = 0.58$), regression equations were generated (Fig. 5). For the sample studied, the regression equations were found to predict nose projection better than the four traditional methods listed above, and estimated values did not differ

TABLE 2a. Predicted nose projection compared to actual nose projection (mm), using methods conducted on entire study sample

	Actual	Gerasimov (1)	Krogman variation (2)	Prokopec (3)	George (4)
Males					
(n = 29)					
X average	30.9	37.2*	29.0	32.3	29.5
SD	4.7	11.6	10.8	6.5	4.2
X error		6.2	−1.9	1.4	−1.5
SD		9.5	7.6	3.9	3.5
Females					
(n = 30)					
X average	28.1	32.4*	24*	30.3*	25.3*
SD	4.1	10.8	8.8	5.1	4.5
X error		4.3	−4.1	2.2	−2.8
SD		8.4	7.0	3.9	2.4

* Indicates statistically significant difference from actual values, $P < 0.01$ (equivalent to $P < 0.05$ after Bonferroni adjustment for five tests).

TABLE 2b. Predicted nose projection compared to actual nose projection (mm) for original method of krogman (method 2) that could only be used on a subset of total study sample

	Actual	Krogman (2)
Male subsample (n = 12)		
X Average	31.1	25.4
SD	5.0	9.2
X Error		−5.7
SD		6.4
Female subsample (n = 14)		
X Average	28.7	24.4
SD	3.8	6.5
X Error		−4.3
SD		6.0

* Indicates statistically significant difference from actual values, $P < 0.05$.

from actual values at statistically significant levels (average error for both males and females = 0.0 ± 2.7 mm, Tables 2 and 3).

It was also found that the ratio of vertical soft-tissue pronasale height (pronasale to nasion (y)) to hard-tissue nasal height (point X to nasion; see Fig. 1) did not differ substantially, being on average 0.82 ± 0.05 in males and 0.84 ± 0.04 in females. Nasal spine angle (as measured from the Frankfurt horizontal) was found to explain some of the variance in the ratio (males $r^2 = 0.26$, females $r^2 = 0.22$), and when regression equations were generated (Fig. 5), they predicted nasal height in this sample rather accurately, with estimated values not differing from actual values at statistically significant levels (average error for males = 0.4 mm, SD 2.3 mm; average error for females = −0.3 mm, SD 2.0 mm; see Table 3). Despite some relationship to soft-tissue nasal height (see above), nasal spine angle was hardly related to the general direction of the columella (male $r^2 = 0.07$, female $r^2 = 0.05$).

When the two regression equations generated for each sex in this study were used to position the pronasale from the skulls in this sample, the average shortest distance between the predicted and actual position (male mean = 2.3 mm, SD 1.7 mm; female mean = 2.6 mm, SD 1.6 mm) was much less and more accurate than that given by the traditional methods (method 1: males = 11.8 mm, SD 9.0 mm, females = 10.5 mm, SD 6.7 mm; method 3: males = 4.6 mm, SD 1.9 mm, females = 4.4 mm, SD 2.2 mm; Table 3).

DISCUSSION

Irrespective of the planes used for measuring, the average sex-specific values of nose projection (x) tended to be ~4 mm smaller in this study in comparison to others. In adults (mean age 17 years), Genecov et al. (1990) found nose projection from nasion to pronasale measured along the Frankfurt horizontal to be about 36 mm, SD 3.5 mm, for males and 34 mm, SD 3.5 mm, for females. In a study by Nanda et al. (1990), the mean projection value for females at age 18 years, as measured from nasion to pronasale, perpendicular to the pterygomaxillary vertical plane, was ~29 mm, while for males it was ~34 mm. Nose height values (y) were approximately 5 mm larger in this study in comparison to those of Nanda et al. (1990), who found nose height (nasion to pronasale, measured parallel to the pterygomaxillary vertical plane) to be ~38 mm in females at age 18, while for males it averaged just over 39 mm. The same sexual dimorphism pattern was found in this study as in others, with males having larger values than females (Genecov et al., 1990; Nanda et al., 1990; Posen, 1967; Subtelny, 1959). Consistent with the findings of others (e.g., Genecov et al., 1990; Nanda et al., 1990), this study found differences between the sexes to be about two times greater for nose projection (x) than nose height (y).

Method 1: Gerasimov technique

This technique was highly unreliable, performing the worst of all four methods, and often resulted in an overestimation of nose projection, as suggested previously (Stephan and Henneberg, 2001). To us, this was not surprising, since the technique seemed rather subjective and imprecise because it relied on the "general directions" of two bones; yet many facial approximation practitioners appear to regard this description as being precise enough for accurate replication (e.g., Fedosyutkin and Nainys, 1993; Gerasimov, 1971; Kustar, 1999; Prag and Neave, 1997). We suggest that this technique not be used to construct noses in future facial approximations.

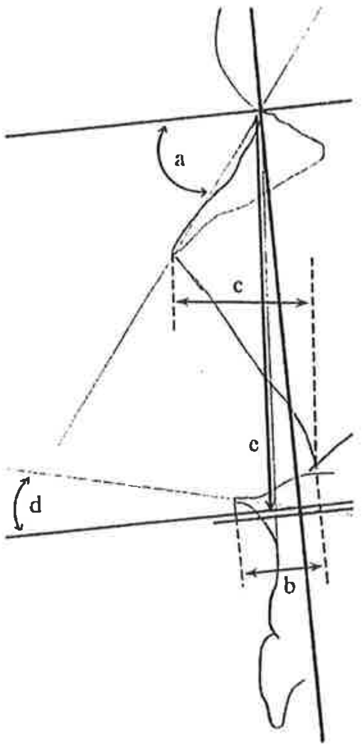
Method 2: Krogman technique

Both the actual technique and the variation reported in this paper performed rather badly, being third best out of four methods. Based on the evidence found in this sample, it seems that these methods should not be used in future facial approximations. However, this recommendation may be harsh because: 1) the determination of the vomer/

TABLE 3. Predicted pronasale position compared to actual pronasale position (mm)

	Actual	Gerasimov (1)	Prokopec (3)	Regression
Males (n = 29)				
X average	30.9	37.2*	32.3	31.1
SD	4.7	11.6	6.5	3.7
X error		6.2	1.4	0.2
SD		9.5	3.9	2.7
Y average	44.2	51.7**	45.8**	44.7
SD	3.5	7.7	3.4	2.8
Y error		7.5	1.6	0.4
SD		6.4	2.6	2.3
Average shortest distance		11.8	4.6	2.3
SD		9.0	1.9	1.7
Females (n = 30)				
X average	28.1	32.4*	30.3	28
SD	4.1	10.8	5.1	3.1
X error		4.3	2.2	−0.1
SD		8.4	3.9	2.6
Y average	43.0	49.7**	43.9**	42.3
SD	3.5	6.7	3.5	3.1
Y error		6.7	1.0	−0.3
SD		4.9	2.1	2
Average shortest distance		10.5	4.4	2.6
SD		6.7	2.2	1.6

* Indicates statistically significant difference from actual values, $P < 0.01$ (equivalent to $P < 0.05$ after Bonferroni adjustment for five tests).
** Indicates statistically significant difference from actual values, $P < 0.017$ (equivalent to $P < 0.05$ after Bonferroni adjustment for three tests).



	r^2	S.E.
Pronasale projection:		
Males $x = -0.32(a) + 0.85(b) - 0.42(c) + 49.58$	0.66	2.80
Females $x = -0.41(a) + 0.37(b) + 49.87$	0.58	2.70
Pronasale height:		
Males and females $y = (-0.002(d) + 0.83)*(e)$	0.64	2.17

Fig. 5. Regression equations generated to predict pronasale position (x and y) in relation to nasion and Frankfurt horizontal. Image illustrates variables used in equations. Black lines indicate those in same plane as Frankfurt horizontal or perpendicular to it. a, nasal bone angle, as measured from nasion to rhinion, from Frankfurt horizontal; b, distance from tip of nasal spine to border of nasal aperture (at its base); c, distance from rhinion to most posterior point on nasal aperture border, measured perpendicular to nasion/prosthion plane (see line A, method 3, Fig. 1); d, nasal spine angle, as measured from Frankfurt horizontal (positive if above Frankfurt horizontal; negative if below); e, distance of point X from nasion (also see Fig. 1, method 4). All linear measurements measured in millimeters at life size (i.e., radiographic measures rescaled).

maxilla junction may be more accurate on real skulls than in radiographs; and 2) the distance to the profile line of the nasal aperture does not indicate/corroborate any inaccuracy of the original guideline, since this measure may not be the same as that to the vomer/maxilla junction.

However, it is worth noting that the variation of method 2 performed better than the original technique, although sample sizes are small for the latter (Table 2a,b). The difference between the original technique and the variation of method 2, although statistically significant ($P < 0.05$), was only 3 mm, which indicates that differences between the two measures of nasal spine length (from the volmer/maxilla junction and from the nasal aperture) were only about 1 mm, since the measures are multiplied by 3. The difference of 1 mm does not appear to be significant, because the error in determining the vomer/maxilla junction on real skulls may be of equal magnitude. For example, the dots Taylor (2001) used to mark the vomer/maxilla junction are ~1 mm in diameter, and measurements from the dots do not appear to be consistent, i.e., not always from the dot's center or its distal/proximal border (Taylor, 2001, p. 393). This seems to indicate that both method 2 and its variation should not be used in future facial approximation methods.

This study demonstrates a significant limitation of the Krogman method because it relies on multiplying the nasal spine length by 3, which magnifies any error by 3×. Furthermore, the use of the midphiltrum soft-tissue depth seems illogical when it is not directly related to the nose and more directly associated depths exist, like that at the subnasale (Helmer, 1984). In addition, placing the midphiltrum depth so that it follows the general direction of the nasal spine, as indicated by Taylor (2001), seems problematic, since the soft-tissue midphiltrum is actually located inferiorly to the hard-tissue midphiltrum (George, 1993). From images provided by Taylor (2001) and Gatliff and Taylor (2001), it can be seen that the midphiltrum depth is often placed at the level of the soft-tissue subnasale, which is unrealistic. Using the midphiltrum depth to represent subnasale is inaccurate, since the subnasale depth appears to be about ~5 mm greater than that reported by Rhine and Moore (1984) at midphiltrum (Helmer, 1984). In Figure 12.25 of Taylor (2001, p. 394), used to specifically describe the determination of nose projection, the midphiltrum marker actually falls well above subnasale, which clearly contradicts the directions of Rhine and Campbell (1980) for the placement of this soft-tissue depth marker.

Other inaccuracies may also be introduced by the use of large, cumbersome rubber cylinders to mark soft-tissue depths, as recommended by Taylor (2001), Gatliff (1984), Gatliff and Snow (1979), and Gatliff and Taylor (2001), since they sometimes cannot be placed flush against the skull at particular anatomical locations, such as the nasion and subnasale (see Taylor, 2001, soft-tissue depth no. 5, midphiltrum, Fig. 11.27, p. 357; and soft-tissue depth no. 3, nasion, Fig. 12.16, p. 387). An alternative that seems more appropriate is to make a precise plaster or acrylic skull cast, and bore holes to locate small-diameter, pointed, soft-tissue markers (e.g., stainless steel nails or sharpened plastic/wood rods) at more realistic angles and exact depths (R. Taylor and C. Stephan C, 2001, personal communication).

Method 3: Prokopec/Ubelaker technique

This technique performed rather well. However, standard deviations of error were higher than those of method 4 for nose projection. While pronasale prediction was considerably more accurate than in method 1, new methods described in this paper performed better. Therefore, this method should not be used if the regression equations described here, or method 4, can be used instead.

Despite the accuracy of this method in predicting pronasale position, observations made in this study suggest that much caution should be used when employing this method to predict profile nose shape, as suggested by Gerasimov (1971) and Prokopec and Ubelaker (2002). However, this is the subject of further study and will not be discussed here.

Method 4: George technique

This method estimated nose projection well, being the best of the four methods studied. The success of this method seems to be attributable to the finding that nose projection did not appear to vary considerably with nasion-point A height. Since this technique was based on methods to predict "aesthetically pleasing" nose projections (see introduction), it seems surprising that the method worked so well. The accuracy of this technique appears to indicate that either natural nose projections are "aesthetic" for the majority of the population, or that the sample studied happened to possess "aesthetic noses" despite its random selection.

General discussion

Two of the four traditional methods of facial approximation used to predict nose projection and/or pronasale position were inaccurate. It seems that new methods reported in this study predict nose dimensions better than the other two traditional methods that were quite accurate (methods 3 and 4). However, these regression equations need to be tested in other samples before they can be considered robust. It thus seems that either the new methods reported here or the methods of George (1987) and Prokopec and Ubelaker (2002) can be used to estimate nose projection and/or pronasale position with a rather high degree of accuracy in future facial approximations; however, further reliability studies are needed.

When estimating nose projection from dry skulls, caution may be needed if the distal ends of the nasal bones have changed shape as a result of dehydration (R. Taylor, 2001, personal communication). We noticed arching of the nasal bones in the coronal plane after processing and drying of skulls with long nasal bones (like those of koalas, possums, and kangaroos), but this phenomenon is yet to be systematically measured and verified in humans. Since human nasal bones are rather short and robust in comparison to those of other mammals, the amount of dehydration-related change may be minimal. However, the possibility that it may occur should not be disregarded. If it does cause significant changes, then the soft-tissue relationships of the hydrated nasal skeletal profile found in this study are probably different from those that exist between hydrated (living) soft tissues and dehydrated (dry) skulls.

The finding that the two most popular methods of determining nose projection/pronasale position are inaccurate suggests that many facial approximations have been made with inaccurate noses. It also suggests that practitioners using various methods will generally produce, on average, noses with different projections, e.g., method 1 leads to further projecting noses than method 2. Some support may be found for this in the literature. For example, profile images of facial approximations by Prag and Neave (1997), who used method 1, typically appear to have slightly more projecting noses than those typically represented on facial approximations by Gatliff (2001), who used method 2. It is worth noting here that frontal images are of little use in evaluating nose projection, because lighting shadows often produce misleading impressions (e.g., see Gatliff, 2001, Fig. 13.48, p. 463). Given that the popular methods of nose projection used in this study lead in some instances to extreme and even unrealistic estimations of nose projection in a sample of 59 individuals, it is expected that some facial approximations in the literature would also display unreal nose projections if guidelines were objectively followed. However, this appears not to be the case, and suggests that either facial approximations with unrealistic noses are not published, or that facial approximation practitioners in some instances "curb" guidelines (maybe unconsciously) when the noses they predict seem unrealistic. If the latter is done, it seems reasonable, given that facial approximation attempts to build a face as accurately as possible. However, one has to wonder why specific nose projection guidelines were ever used in the first place, if they require subjective adjustments to be made ad libitum?

Findings that previously untested, subjective methods of facial approximation are inaccurate seem to be a common research result. Traditional methods used to predict ear height (Farkas et al., 1987), eyeball projection (Stephan, 2002c), mouth width (Stephan, 2002b), and position of the superciliare (Stephan, 2002d) have also been shown to be unreliable when tested. In light of these findings, it does not seem surprising that facial approximation may be an inaccurate technique that rarely achieves its objective of specific and purposeful facial identification of the target individual, as was suggested (Montagu, 1947; Stephan and Henneberg, 2001). These findings indicate that facial approximation may achieve success through methods other than facial recognition, such as contextual information (Haglund and Reay, 1991) or chance (Stephan and Henneberg, 2001). It may be that facial approximation acts to promote public interest in the case, which in turn generates extra leads through the mechanisms suggested above, without the facial approximation being specifically recognized during critical stages (Stephan, 2002b; Ubelaker, 1993). It is also important to note that people probably *think* they recognize the target individual from the facial approximation *after* they know who the target individual is. It therefore seems that despite the inaccuracy of the facial approximation technique, it is useful to forensic work, since it generates public interest. However, it should be noted that successes due to specific and purposeful facial recognition are probably unlikely (Stephan, 2002b; but for case-related evidence, see Ubelaker, 1993).

While this study systematically examines one aspect of soft-tissue nose prediction from the skull, there is much nose anatomy that, as yet, cannot be predicted with reliable estimates of error, e.g., prediction of the shape of the profile line of the nose, nostrils, nose apex or bulb, columella, and alars. In the context of the entire face, it is clear that facial approximation is based on only a small number of known relationships, which are probably inadequate for building a complete face that is representative of the person to whom the skull belonged. Much work is needed in the future if a comprehensive understanding of the natural soft- to hard-tissue relationships is to be achieved, and if faces built from skulls are to closely represent target individuals.

With further systematic studies of the relationship of the soft tissues to the skull, facial approximation is likely to come to include many technical soft-tissue prediction guidelines. The regression equations presented here are examples of this, as are equations presented by Simpson and Henneberg (2002) that attempt to individualize average soft-tissue measurements based on skull sizes. It seems probable that this will have ramifications for the way facial approximation is approached in the future, because some measurements are not easily taken from skulls directly (e.g., nasal bone angle, which may require planes to travel through the nasal bones if they are medially convex; see Fig. 5), and many calculations will need to be made. Consequently, three-dimensional (3D) computer techniques will probably become the method of choice, because they enable: easily made indirect measurements (i.e., measures that pass through bone planes); the calculation of many variables in a very short time; and the generation of facial approximations in 3D. Therefore, there appears to be a need for facial approximation practitioners to move away from traditional three- and two-dimensional techniques and for 3D computer techniques to be further developed. Unlike 3D clay methods that use actual skulls (or casts), two-dimensional drawing methods, like 3D computer methods, allow some measures to be indirectly made on the skull and are therefore advantaged. However, drawing methods appear to be limited because they are two-dimensional. Despite the disadvantages of the 3D clay facial approximation method mentioned above, this method may continue to play a useful role, since it produces detailed, lifelike three-dimensional "hard copy" models, which may be favored for display, particularly in museums.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Ronn Taylor (University of Melbourne). Over the last several years he has helped mentor C.N.S., and parts of this paper are a direct result of discussions they have had. Thanks go to Tim Anson (University of Adelaide) for permitting the use of skeletal material from the St. Mary's Cemetery skeletal collection for Figure 3. Thanks also go to Robert George (Florida International University) for helpful correspondence, and to Suzanne Langlois and Jack Miniszewski (Royal Adelaide Hospital) for technical assistance and the generation of x-rays used in Figure 3.

LITERATURE CITED

Bruce V. 1988. Recognising faces. Hove: Lawrence Erlbaum Associates.

Brues AM. 1958. Identification of skeletal remains. J Crim Law Criminol Police Sci 48:551—556.

Chaconas SJ. 1969. A statistical evaluation of nasal growth. Am J Orthod 56:403–414.

Cherry DG, Angel JL. 1977. Personality reconstruction from unidentified remains. FBI Law Enforce Bull 46:12–15.

Farkas LG, Munro IR, Kolar JC. 1987. The validity of neoclassical facial proportion canons. In: Farkas LG, Munro IR, editors. Anthropometric facial proportions in medicine. Springfield, IL: Charles C. Thomas. p 57–66.

Farrar F. 1977. From skull to visage: a forensics technique for facial restoration. Police Chief 44:78–80.

Fedosyutkin BA, Nainys JV. 1993. The relationship of skull morphology to facial features. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull. New York: Wiley-Liss. p 199–213.

Fraser I, Parker D. 1986. Reaction time measures of feature saliency in a perceptual integration task. In: Ellis H, Jeeres M, Newcombe F, Young A, editors. Aspects of face processing. Dordrecht: Martinus Nijhoft. Q: 2

Gatliff BP. 1984. Facial sculpture on the skull for identification. Am J Forensic Med Pathol 5:327–332.

Gatliff BP, Snow CC. 1979. From skull to visage. J Biocommun 6:27–30.

Gatliff BP, Taylor KT. 2001. Three-dimensional facial reconstruction on the skull. In: Taylor KT, editor. Forensic art and illustration. Boca Raton: CRC Press. p 419–475.

Genecov JS, Sinclair PM, Dechow PC. 1990. Development of the nose and soft tissue profile. Angle Orthod 60:191–198.

George RM. 1987. The lateral craniographic method of facial reconstruction. J Forensic Sci 32:1305–1330.

George RM. 1993. Anatomical and artistic guidelines for forensic facial reconstruction. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull. New York: Wiley-Liss. Q: 3

Gerasimov M. 1971. The face finder. London: Hutchinson & Co.

Haglund WD, Reay DT. 1991. Use of facial approximation techniques in identification of Green River serial murder victims. Am J Forensic Med Pathol 12:132–142.

Haig ND. 1984. The effect of feature displacement on face recognition. Perception 13:505–512.

Haig ND. 1986. Exploring recognition with interchanged facial features. Perception 15:235–247.

Helmer R. 1984. Schadelidentifizierung durch elekronicshe Bildmischung: Sugl el Beitr zur Konstitutionsbiometrie U Dickermessung d Gesichtsweichteile. Heidelberg: Krminalistik-Verlag. AQ: 4

Helmer RP, Rohricht S, Petersen D, Mohr F. 1993. Assessment of the reliability of facial reconstruction. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull. New York: Wiley-Liss. p 229–246.

Krogman WM. 1946. The reconstruction of the living head from the skull. FBI Law Enforce Bull 17:11–17.

Krogman WM. 1962. The human skeleton in forensic medicine. Springfield, IL: Charles C. Thomas.

Kustar A. 1999. Facial reconstruction of an artificially distorted skull of the 4th to the 5th century from the site of Mozs. Int J Osteoarchaeol 9:325–332.

Macho GA. 1986. An appraisal of plastic reconstruction of the external nose. J Forensic Sci 31:1391–1403.

Montagu MFA. 1947. A study of man embracing error. Technol Rev 49:345–347.

Nanda RS, Meng H, Kapila S, Goorhuis J. 1990. Growth changes in the soft tissue facial profile. Angle Orthod 60:177–190.

Perper JA, Patterson GT, Backner JS. 1988. Face imaging reconstructive morphography: a new method for physiognomic reconstruction. Am J Forensic Med Pathol 9:126–138.

Phillips VM, Rosendorff S, Scholtz HJ. 1996. Identification of a suicide victim by facial reconstruction. J Forensic Sci 14:34–38.

Posen J. 1967. Longitudinal study of growth of the nose. Am J Orthod 53:746–756.

Powell N, Humphreys B. 1984. Proportions of the aesthetic face. New York: Thieme-Stratton.

Prag J, Neave R. 1997. Making faces: using forensic and archaeological evidence. London: British Museum Press.

Prokopec M, Ubelaker DH. 2002. Reconstructing the shape of the nose according to the skull. Forensic Sci Commun 4. AQ: 5

Rathbun TA. 1984. Personal identification: facial reproductions. In: Rathbun T, Buikstra J, editors. Human identification: case studies in forensic anthropology. Springfield, IL: Charles C. Thomas. p 347–362.

Rhine JS, Campbell HR. 1980. Thickness of facial tissues in American blacks. J Forensic Sci 25:847–858.

Rhine JS, Moore CE. 1984. Tables of facial tissue thickness of American Caucasoids in forensic anthropology. Maxwell Museum Technical Series 1. AQ: 6

Simpson E, Henneberg M. 2002. Variation in soft tissue thicknesses on the human face and their relation to craniometric dimensions. Am J Phys Anthropol 118:121–133.

Snow CC, Gatliff BP, McWilliams KR. 1970. Reconstruction of facial features from the skull: an evaluation of its usefulness in forensic anthropology. Am J Phys Anthropol 33:221–228.

Stephan C. 2002a. Do resemblance ratings measure the accuracy of facial approximations. J Forensic Sci 47:239–243.

Stephan CN. 2002b. Facial approximation: an evaluation of mouth width determination. Am J Phys Anthropol (in press). AQ: 7

Stephan CN. 2002c. Facial approximation: falsification of globe projection guideline by exophthalmometry literature. J Forensic Sci 47:1–6.

Stephan CN. 2002d. Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline. Forensic Sci Int (in press). AQ: 8

Stephan C, Henneberg M. 2001. Building faces from dry skulls: are they recognized above chance rates? J Forensic Sci 46:432–440.

Stoney MB, Koelmeyer TD. 1999. Facial reconstruction: a case report and review of development of techniques. Med Sci Law 39:49–60.

Subtelny JD. 1959. A longitudinal study of soft tissue facial structures and their profile characteristics, defined in relation to underlying skeletal structures. Am J Orthod 45:481–507.

Suk V. 1935. Fallacies of anthropological identifications. Publ Facultae Sci Univ Masaryk 207:3–18.

Suzuki T. 1973. Reconsititution of a skull. Int Crim Police Rev 264:76–80.

Taylor KT. 2001. Forensic art and illustration. Boca Raton: CRC Press.

Ubelaker DH. 1993. Facial reconstruction [author's response]. J Forensic Sci 37:1442–1444.

van Rensburg J. 1993. Accuracy of recognition of 3-dimensional plastic reconstruction of faces from skulls. Anatomical Society of Southern Africa. p 20. AQ: 9

Virchow H. 1912. Die anthropologische Untersuchung der Nase. Z Ethnol 44:290–337.

Virchow H. 1924. Zur Anthropologie der Nase. Z Ethnol 56:94–121.

Appendix 5: Eyeball Projection Paper, *J Forensic Sci*

Stephan, C. N. (2002). Facial approximation: globe projection guideline falsified by exophthalmometry literature. *Journal of Forensic Sciences*, *47*(4), 730-735.

Appendix 6: Superciliare Position Paper, *Forensic Sci Int*

ELSEVIER

Forensic Science International 130 (2002) 29–33

Forensic Science International

www.elsevier.com/locate/forsciint

Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline

C.N. Stephan*

Department of Anatomical Sciences, Adelaide University, 5005 Adelaide, Australia

Received 1 April 2002; received in revised form 3 July 2002; accepted 23 July 2002

Abstract

It has been suggested in the literature that superciliare is located directly above the most lateral point of the iris and that this association may be of use in facial approximation. However, the relationship between the lateral iris and superciliare has not been tested and its accuracy remains unknown. This study aims to determine the accuracy of this relationship using metric and non-metric analysis. The horizontal distance from superciliare to the lateral iris was measured, using photogrammetric methods, in Australians of European extraction (27 males, 48 females), central/south-east Asian extraction (20 males, 19 females) and individuals from other population groups (7 males, 7 females). Results indicate that superciliare position is best approximated by the lateral iris in females. On average, superciliare fell lateral to the lateral iris by 4.8 mm, S.D. 3.4 mm in males, and 1.2 mm, S.D. 5.4 mm in females. In approximately 70–80% of the sample, the superciliare fell between the exocanthion and the pupil center on both sides. It is suggested that the proposed guideline that the lateral iris is equal to superciliare is not very accurate, especially for males. Also the large standard deviations indicate that the position of superciliare is highly variable. The above measures should, on average, give a more accurate prediction of superciliare in contrast to the lateral iris border, and therefore, they should be used in facial approximation. However, the large variation in superciliare position should be acknowledged.

Keywords: Forensic science; Facial reproduction; Facial reconstruction; Eyebrow; Face; Photogrammetry

1. Introduction

Guidelines that predict particular facial features from other pre-existing, or pre-determined, facial features are useful to various disciples where parts of the face may need to be built accurately, such as forensic facial approximation or reconstructive surgery.

Many soft tissue prediction guidelines have been proposed, for building faces from skulls (facial approximation) however, most guidelines have not been scientifically tested and published. Consequently, the accuracy and reliability of these guidelines remain unknown, like for nose projection being equal to three times the length of the nasal spine [1,2] or equal to the junction of tangents following the nasal spine and the last one-third of the nasal bones [3–6]. It is unlikely that these subjectively determined guidelines reliably account for individual variation because they have not been based on any empirical evidence.

Farkas et al. [7] have shown that the length of the ear is not equal to the height of the nose since 90% of people have an ear larger than their nose. It has also been shown that ear inclination is not equal to that of the nose [8]. Stephan [9] has shown that traditional guidelines used in facial approximation for determining mouth width from canine width and interpupillary distance are inaccurate, as is the guideline used to determine globe projection from mid-superior and mid-inferior orbital rims [10].

Since it appears that subjectively determined guidelines display rather large inaccuracies when tested, the verification of other subjective/artistic guidelines is a logical step. It is also necessary if the reliability and accuracy of facial approximation methods are to be determined and possibly improved upon.

One guideline that has been suggested for facial approximation is that superciliare (the most superior part of the

---

* Tel.: +61-8-8303-6326; fax: +61-8-8303-4398.
E-mail address: carl.stephan@adelaide.edu.au (C.N. Stephan).

0379-0738/02/$ – see front matter © 2002 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0379-0738(02)00279-7

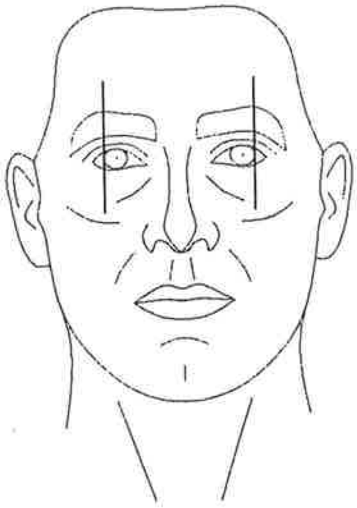


Fig. 1. Guideline for determining superciliare from the lateral iris.

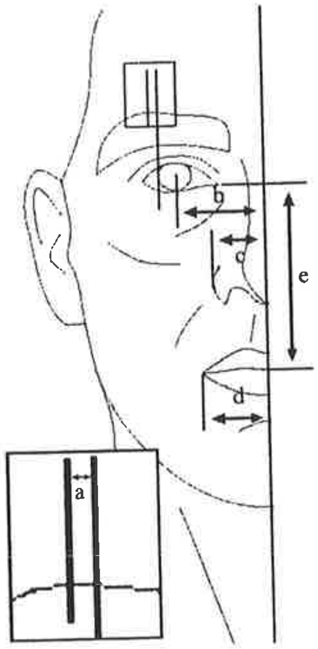


Fig. 2. Bilateral measures taken in this study: (a) distance of superciliare from lateral iris border, "a" was positive if it fell lateral to the lateral iris and negative if it fell medial; (b) the distance from the midline to the pupil center; (c) the distance from midline to alare; (d) the distance from midline to cheilion; (e) the vertical distance between the endocanthion and stomion.

eyebrow) is located directly above the lateral point of the iris [11] (Fig. 1). This guideline has also been used in the past to cosmetically position the arch of the female eyebrow when plucking or waxing [12,13] and a similar guideline has been used in plastic surgery with superciliare being aligned based on the lateral limbus, which lies near the lateral iris [14–16].

The guideline of superciliare being located above the most lateral point of the iris may be close to the truth since the arch of the eyebrow generally appears directly above the eye. This study is aimed at evaluating the accuracy and reliability of this guideline, both metrically and non-metrically, using photogrammetric methods.

2. Materials and methods

One hundred and twenty-eight participants, aged 18–30 years, average 21.4 years, S.D. 3.8 years were photographed at the School of Dental Science, The University of Melbourne, on a specialized craniofacial rig that uses a projected light range finding system to maintain consistent subject-camera distances [17,18]. Frontal photographs of participants (in a relaxed, natural head position, with lips closed) were taken using a Nikon FM-2 35 mm camera fitted with a Nikon 105 mm 1:2.8 Macro Nikor lens, at a distance of 1204 mm (from the film plane to glabella of the subject). Subjects were illuminated by two Elinchrom Prolinca 2500 self-contained studio flash units, positioned higher than the subject (for more information on the photography set-up, see [17,18]). Photographs were scanned into a computer using a Nikon® SF-2000 slide scanner and measured in Adobe® Photoshop® 6.0. The resultant pictures were 1200 pixels in width, and 1803 pixels in height.

All images were rotated as required so that the midsaggital plane, as defined by Farkas [19], was exactly vertical. Superciliare was defined as the most readily determinable superior point of the eyebrow when all of the face could be seen using a Diamond View® 1995SL 483 mm monitor. The horizontal distance from the lateral iris to superciliare was measured on each side of the face (Fig. 2). Four other measures were taken to determine if there was any relationship between them and the position of superciliare. Those measures were: the distance from the midline to the pupil center; the distance from midline to alare; the distance from midline to cheilion; and the vertical distance between the endocanthion and stomion (Fig. 2). All measures were adjusted by a scaling factor of 4.975124 (as determined by photographing an object of known length also at 1204 mm from the camera) to obtain actual values in millimeters [18].

Participants were also asked to indicate if they plucked/waxed their eyebrows and in which location they did so according to Fig. 3. Although plucking in region 2 has the potential to dramatically alter the position of superciliare the decision was made not to exclude these individuals since they may also be subject of a forensic inquiry.

F-tests and histograms were used to compare data before the use of the relevant *t*-test (either equal or unequal variance). Significance was initially set at the 95% confidence level, but altered according to the Bonferroni adjustment. Pearson's correlations were also determined between the

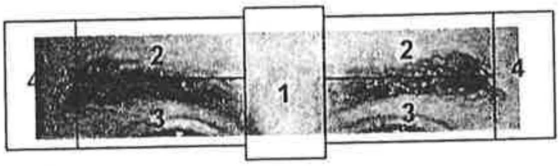


Fig. 3. Illustration of the zones used to describe eyebrow plucking/waxing.

position of superciliare from the lateral iris and the other measurements.

3. Results

Overall, superciliare was found to be approximately 2.7 mm lateral to the lateral iris. However, values showed a large distribution resulting in high standard deviations (5.0 mm). Table 1 summarizes the results for each sample. No statistically significant differences in horizontal superciliare location were detected between the sides or the populations of origin. There were, however, statistical significant differences between the sexes for all populations ($P < 0.017$). On average, male superciliares were 3.5 mm more lateral to the iris than in females and had smaller standard deviations (Table 1). Non-metrically, 80% of right superciliares could be found in the region between the pupil center and the exocanthion. For the left superciliare, almost 70% could be found between the pupil center and the exocanthion.

Other measures showed weak correlations with superciliare, however, actual superciliare values were disparate and therefore attempts at its prediction were unsuccessful. As a result, this aspect of the study will not be further discussed.

Overall, few males reported eyebrow manipulation. Of those that did (15% of male Europeans and 5% of Asians) most removed hair in zone 1 (see Fig. 3). Sixty-five percent of female Europeans removed eyebrow hair, 23% of this group removing hair in zone 2. Eighty-four percent of female Asians removed eyebrow hair, 47% removing hair from zone 2. The female European sample was the only group large enough to test for differences between those that removed portions of their eyebrow and those that did not. No statistically significant results were found between these groups, although averages were closer to the lateral iris for those females who removed eyebrow hair (Right side = 0.13 mm, Left side = 0.32 mm) in comparison to those that did (Right side = 1.79 mm, Left side = 2.02 mm). Standard deviations were also generally greater for the hair removal group (Right side = 4.88 mm, Left side = 6.96 mm) than the group that did not remove hair (Right side = 4.65 mm, Left side = 4.84 mm).

Table 1
Summary table of horizontal distance from superciliare to the lateral iris for males and females separated by population of origin

	n	Right side (mm)		Left side (mm)	
		Average	S.D.	Average	S.D.
Male European	27	4.4	2.7	4.2	3.7
Female European	48	0.7	4.8	0.9	6.3
Male Asian	20	4.9	4.5	5.8	2.5
Female Asian	19	2.3	5.2	2.5	5.1
Other male individuals	7	6.4	3.6	4.1	3.0
Other female individuals	7	0.4	4.4	−0.8	2.8
Total male	54	4.8	3.5	4.8	3.2
Total female	74	1.1	4.9	1.2	5.8
Total	128	2.7	4.7	2.7	5.2

4. Discussion

The guideline that superciliare is located directly above the lateral iris may appear to be fairly accurate, especially in females, since small differences were found between the horizontal position of superciliare and the lateral point of the iris (1 mm for females and 5 mm for males). However, the large variation in superciliare position, as reflected by large standard deviations (especially for females), and the observation that up to 30% of people's superciliares fall outside of the region between their exocanthion and the pupil center demonstrate the large inaccuracy of this guideline and its limited usefulness. It is suggested, therefore, that the guideline that the horizontal displacement of superciliare is equal to the lateral iris, is only used in a very general sense, if at all, in future facial approximation methods.

Since the lateral limbus of the sclera/cornea could not be precisely located from frontal photographs, the accuracy in using this landmark to predict the horizontal position of superciliare could not be evaluated. Since the limbus may fall slightly lateral to the lateral border of the iris, this guideline maybe slightly more accurate than using the lateral iris border itself. However, the magnitude of the measurements found for males in this study would suggest that this guideline is also inaccurate, at least for predicting male superciliares. Overall, it seems that this guideline should also only be used in a very general way to predict superciliare position and practitioners should note the wide range of variability in actual superciliare location.

It appears that the finding of a large proportion (up to 80% in this sample) of people's superciliares falling between the region of the exocanthion and the pupil center will be useful to facial approximation practitioners. The finding that superciliare is, on average, 2 mm lateral to the lateral iris, and that this distance increases for males will probably also be useful.

It should be noted that some difficulty was encountered in the placing of superciliare due to the structure of the eyebrows, which are not generally well-defined arches. The individual hair fibers in the superior mid-portion of the

Fig. 4. Example of male eyebrows (a) smooth density transition to the "main brow"; (b) abrupt density transition to the "main brow". Black arrows indicate region of superciliare and white arrow shows region of "false superciliare".

eyebrow are angled inferiorly and laterally, becoming denser toward the center of the brow, but the density transition may be smooth or abrupt (Fig. 4). In smooth brows it is likely, from a distance, that the judgment of the superior part of the brow arch would be lower than it is in reality due to the scarcity of superior hair follicles. This placement of superciliare may be higher, from a distance, than the relative placement of a "false superciliare" in a person who has an abrupt density change (Fig. 4). Also, technically defining superciliare appears to be generally quite difficult since eyebrow hairs may be present, although sparse, surprisingly high up on the forehead. Therefore, the determination of superciliare depends on the subjective interpretation of where hair density justifies it (Fig. 4 illustrates where superciliare was positioned in this study). The approach was taken in this study to define superciliare as being the highest point on the eyebrow that is readily determinable when the whole face was in view on the computer monitor (however, as seen in Fig. 4, numerous other hairs were present above and lateral to the position of superciliare).

It is necessary to acknowledge that photogrammetric methods are limited by magnification and perspective distortions inherent to all photography methods [20–23]. However, measures of superciliare to the lateral iris border in this study are expected to be affected little since a large focal length lens was used, measures of superciliare and lateral iris borders were close to the camera focal plane at glabella, and the distance being measured was small (so absolute error in comparison to larger measures was much less). The validity of photogrammetric measures of the eyes is also supported by other studies that have found these measures not to differ from direct anthropometric measures in living subjects [21,22].

Although additional suggestions have been made for determining other features of the eyebrow it appears that these should be regarded with some caution until they have been tested and verified. These guidelines include: the medial most point of the brow falls vertically inline with the alare [14,15]; the lateral most point of the brow falls inline with a tangent connecting alare to the exocanthion [14,15]; the medial and lateral brow should fall horizontal to each other [14,15]; in males the brow arc is at the supra-orbital rim and in females it is above it [14]; individuals with strongly developed supra-orbital margins have lower brows [3]; strongly developed supra-orbital margins and brow ridges indicate an acute angle of the brow arch where less developed supra-orbital margins and brow ridges indicate a more smoothly arched brow [3].

In contrast to the specific focus of this paper, Rozprym [24] has presented a more general study of eyebrow morphology (and eyelashes) in over 500 individuals. Like the findings of the current study, Rozprym [24] found large variations in eyebrow form between individuals. The eyebrow categorizations Rozprym [24] proposed also appear to be useful, however, since this study was conducted over 75 years ago, further studies that aim to test the repeatability of Rozprym's categories and character frequencies on modern samples would appear to be useful.

Oestreicher and Hurwitz [25] have also conducted studies on the position of the eyebrow, but did not specifically address superciliare. Oestreicher and Hurwitz [25] studied the right eyebrow on 46 males and 30 females. Their results indicated that although the mid-eyebrow height tended to decrease with age, with respect to the superior orbital rim, the trend was not statistically significant ($P > 0.05$). The difference between the young (<40 years) and old groups (>59 years) was only 0.71 mm. However, females were found to have eyebrows, which were higher than males at statistically significant levels (males = -3.23 mm, females = -0.99 mm, "–" indicates brow position below the superior orbital rim, $P < 0.004$).

The results of this study are consistent with the results of others that indicate that current methods used in facial approximation are often inaccurate [7,9,10]. Consequently, the inaccuracies in the method may make it unlikely that facial approximations can reliably achieve purposeful and specific recognition. This lends support to the theory that facial approximation success may be due to factors other than facial recognition itself [26,27].

In summary, the guideline to determine superciliare from the lateral iris was shown in this study to be unreliable. Large variations in superciliare position were observed with approximately 70–80% of people's superciliares ($N = 128$) falling within the region between the exocanthion and the pupil center on both sides. On average, superciliare fell 2.6 mm lateral to the lateral iris, however, statistically significant differences were found between males and females, and large standard deviations were observed.

Repeatability studies and further tests of facial approximation guidelines are needed in the future to help assess the accuracy of the facial approximation method and improve it.

Acknowledgements

Thanks to my Ph.D. supervisors Professor Maciej Henneberg (The University of Adelaide) and Professor John Clement (The University of Melbourne) who have provided me with much support in all my work. Thanks also to Professor David Perrett (The University of St. Andrews) for making Psychomorph Software available, on a continuing basis, enabling parts of Figs. 1 and 2 to be created. Thanks also to the School of Dental Science, The University of Melbourne, for use of photography facilities and to the Prince Alfred Hotel, Melbourne, for its support toward participant recruitment.

References

[1] W.M. Krogman, M.Y. İşcan, The Human Skeleton in Forensic Medicine, Charles C. Thomas, Illinois, 1986.

[2] B.P. Gatliff, Facial sculpture on the skull for identification, Am. J. Forensic Med. Pathol. 5 (1984) 327–332.

[3] B.A. Fedosyutkin, J.V. Nainys, The relationship of skull morphology to facial features, in: M.Y. İşcan, R.P. Helmer-Book (Eds.), Forensic Analysis of the Skull, Wiley-Liss, New York, 1993, pp. 199–213.

[4] A. Kustar, Facial reconstruction of an artificially distorted skull of the 4th to the 5th century from the site of Mozs, Int. J. Osteoarchaeol. 9 (1999) 325–332.

[5] J. Prag, R. Neave, Making Faces: Using Forensic and Archaeological Evidence, British Museum Press, London, 1997.

[6] M. Gerasimov, The Face Finder, Hutchinson & Co., London, 1971.

[7] L.G. Farkas, I.R. Munro, J.C. Kolar, The validity of neoclassical facial proportion canons, in: L.G. Farkas, I.R. MunroBook (Eds.), Anthropometric Facial Proportions in Medicine, Charles C. Thomas, Illinois, 1987, pp. 57–66.

[8] M.S. Skiles, P. Randall, The aesthetics of ear placement: an experimental study, Plast. Reconstr. Surg. 72 (1983) 133–140.

[9] C.N. Stephan, Facial approximation: an evaluation of mouth width determination, Am. J. Phys. Anthropol. (2002), in press.

[10] C.N. Stephan, Facial approximation: falsification of globe projection guideline by exophthalmometry literature, J. Forensic Sci. 47 (2002) 1–6.

[11] K.T. Taylor, Forensic Art and Illustration, CRC Press, Boca Raton, 2001.

[12] J. Campsie, Maire Claire: Health and Beauty, Murdoch Books, Sydney, 1997.

[13] J. Campsie, Maire Claire: Hair and Make-up, Murdoch Books, Sydney, 1998.

[14] H.G. Brennan, Correction of the ptotic brow, Otolaryngol. Clin. N. Am. 13 (1980) 265–273.

[15] R. Ellenbogen, Transcoronal eyebrow lift with concomitant upper blepharoplasty, Plast. Reconstr. Surg. 71 (1983) 490–499.

[16] N. Powell, B. Humphreys, Proportions of the Aesthetic Face, Thieme-Stratton, New York, 1984.

[17] T. Dobrostanski, C.D. Owen, Craniofacial photography in the living, in: J.G. Clement, D.L. Ranson (Eds.), Craniofacial Identification in Forensic Medicine, Oxford University Press, New York, 1998, pp. 137–149.

[18] C.N. Stephan, J.G. Clement, C.D. Owen, T. Dobrostanski, A. Owen, A new rig for craniofacial photography put to the test, 2002, submitted for publication.

[19] L.G. Farkas, Sources of error in anthropometry and anthroposcopy, in: L.G. Farkas (Ed.), Anthropometry of the Head and Face, Raven Press, New York, 1994, pp. 57–70.

[20] J.A. Gavan, S.L. Washburn, P.H. Lewis, Photography: an anthropometric tool, Am. J. Phys. Anthropol. 10 (1952) 331–353.

[21] L.G. Farkas, Photogrammetry of the face, in: L.G. Farkas (Ed.), Anthropometry of the Head and Face, Raven Press, New York, 1994, pp. 79–88.

[22] L.G. Farkas, W. Bryson, J. Klotz, Is photogrammetry of the face reliable? Plast. Reconstr. Surg. 66 (1980) 346–355.

[23] M.Y. İşcan, Introduction of techniques for photographic comparison: potential and problems, in: M.Y. İşcan, R.P. Helmer (Eds.), Forensic Analysis of the Skull, Wiley-Liss, New York, 1993, pp. 57–70.

[24] F. Rozprym, Eyebrows and eyelashes in man: their different forms, pigmentation and heredity, J. R. Anthropol. Inst. 64 (1934) 353–395.

[25] J.H. Oestreicher, J.J. Hurwitz, The position of the eyebrow, Ophthalmic Surg. 21 (1990) 245–249.

[26] C. Stephan, M. Henneberg, Building faces from dry skulls: are they recognized above chance rates? J. Forensic Sci. 46 (2001) 432–440.

[27] W.D. Haglund, D.T. Reay, Use of facial approximation techniques in identification of Green River serial murder victims, Am. J. Forensic Med. Pathol. 12 (1991) 132–142.

Appendix 7: Frontal Delineation Map Landmarks

Table of landmarks for the frontal delineation map (Fig. 32). Midpoint landmarks and left side bilateral landmarks are not shown. Standard anthropological landmarks are shown in bold type.

Label	Soft Tissue Landmark
1	**Right center of pupil (p)**
3	**midline point of inferior border of upper lip (in rest point 3 & 86 form the stomion (sto))**
4	most superior point on iris border
6	most lateral point on iris border
8	most inferior point on iris border
10	most medial point on iris border
20	**endocanthion (en)**
21	point at flexion of medial lower lid
22	**palpebrale inferius (pi)**
24	**exocanthion (ex)**
26	**palpebrale superius (ps)**
36	most medial point on epicanthal fold
38	most superior point on epicanthal fold
40	most lateral point on epicanthal fold
46	most superior point on bridge of nose
49	point of flexion of nose bridge and bulb
50	lateral point encompassing bulb of nose
56	most superior point of alare insertion to the face
57	**Alare (al)**
58	**Subalare (sbal)**
59	point at most superior flexion of nostril
60	base of visible nostril
62	base of bulb of nose
69	point at upper level of right philtrum ridge
70	**crista philtri landmark (cph)**
73	**cheilion (ch)**
75	point on vermillion border below pt.70
76	**labriale superius (ls)**
81	point at flexion of the right lower border of upper lip near the tubercule
86	**midline point of superior border of lower lip (in rest point 3 & 86 form the stomion (sto))**
91	**labriale inferius (li)**
94	point at flexion of antitragus
95	point at flexion of concha
96	point at the superior visible end of the concha
97	point at the visible inferior end of the proximal helix border
99	point of flexion of proximal helix border
100	point at the most inferior proximal aspect of helix border
108	point at most medial inferior aspect of brow
109	point at most medial superior aspect of brow
111	**superciliare (sci)**
113	most lateral point of brow
124	upper point following nasal bridge line

126	lower point following nasal bridge line
130	medial point following cheek prominence
132	lateral point following cheek prominence
136	upper point of nasolabial line
137	midpoint of nasolabial line
138	lower point of nasolabial line
142	right lateral point on mental labial ridge
143	**sublabiale (sl)**
145	superior sagittal point on mental cleft
146	inferior sagittal point on mental cleft
147	**gnathion approximation (gn)**
164	**trichion (tr)**
170	**otobasion superius (obs)**
172	superior junction of tragus and face
173	lateral most edge of tragus
174	inferior junction of tragus and face
176	**otobasion inferius (obi)**
177	**gonion approximation (go)**
181	**superaurale (sa)**
184	point opposite 95 on free border of ear
186	**subaurale (sba)**
197	point opposite 177 on free border of neck
198	point opposite, but just inferior to 178, on neck
200	point at inferior region of visible neck
201	point on inferior sternocleidomastoid line
203	point on superior sternocleidomastoid line
207	point opposite 197
208	point opposite 186
209	point opposite 181
213	mid sagittal point above 164

Appendix 8: Profile Delineation Map Landmarks

Table of landmarks for the profile delineation map (Fig. 33). Midpoint landmarks are not shown. Standard anthropological landmarks are shown in bold type.

1	**Otobasion inferius (obi)**
2	**Exocanthion (ex)**
4	point at corneal junction with lower lid
5	**palpebrale inferius (pi)**
8	point at corneal junction with upper lid
9	**palpebrale superius (ps)**
10	most lateral point of epicanthal fold
13	most anterior point of epicanthal fold
15	most anterior point of corneum
17	point at superior margin of iris
18	point at mid margin of iris
19	point at inferior margin of iris
20	**right chelion (ch)**
22	most anterior point on upper vermillion border = **labriale superius approximation (ls)**
25	point at junction of upper incisors and upper lip or **stomion (sto) when relaxed**
26	point at flexion of upper lip curve
27	lower point of anterior incisor or **stomion (sto) when relaxed**
30	point at junction of lower incisors and lower lip or **stomion (sto) when relaxed**
33	most anterior point on lower vermillion border = labriale inferius approximation (li)
36	most anterior point of visible nostril
37	point at most superior flexion of nostril
38	most posterior point of visible nostril
39	**subalare (sbal)**
40	**alar curvature point (ac)**
41	most superior point of alare insertion to the face
42	point on intertragal notch near the juction of the inferior tragus with the face
43	deepest point of intertragal notch
45	point of flexion on antitragus
46	point at flexion between antitragus and antihelix
50	superior point on concha border near the anterior antihelix
51	point on helix-tragal junction directly above the external auditory meatus
52	visible point at the outer helix near concha junction
53	most inferior visible point on inner helix border
54	point on inner helix border opposite 47
56	point on inner helix border opposite 86
57	point on inner helix border directly below on 87
59	point near the proximal anterior border of helix
60	point at flexion of inner helix border near helix-tragal junction
61	visible point at the inner helix near concha junction
62	lateral point of brow
64	point just lateral to superciliare
65	point just medial to superciliare
67	inferior-anterior point of brow
68	**approximation of orbital superios (os)**
70	superior point on nasal bridge line
71	midpoint on nasal bridge line

72	inferior point on nasal bridge line
73	anterior point following cheek prominence
74	midpoint following cheek prominence
75	posterior point following cheek prominence
76	point on upper nasolabial line, near pt. 39
78	point on nasolabial line , near pt. 20
79	inferior point on nasolabial line
80	posterior point on mental labial ridge
82	**subaurale (sba)**
83	point at tip of ear lobe
85	point on free margin of ear opposite pt. 47
86	point near postaurale (pa) so that contour follows free margin of ear
87	**superaurale (sa)**
89	**otobasion superius (obs)**
90	**tragion (t)**
91	point at posterior flexion of tragus
92	point at junction of the lower tragus with the face
94	**gonion approximation (go)**
97	**gnathion approximation (ga)**
99	**pogonion (pg)**
101	**sublabiale (sl)**
104	**subnasale (sn)**
107	**pronasale (prn)**
109	point of flexion of nose bridge and bulb
112	**sellion (se)**
113	point at junction of lower brow and profile of face
114	anterior point of eye brow
115	point at juction of upper brow and profile of face
116	**approximation of glabella (g)**
121	**trichion (tr)**
122	point at flexion of receding hairline (if present)
123	point at flexion of hairline near temple
129	point at flexion of the neck and jaw
130	point at flexion of the neck and Adam's apple
131	point at most projecting point of Adam's apple
132	point at lower border of Adam's apple
133	approximation of suprestenale
134	point on upper sternocleidomastoid line
135	mid point on sternocleidomastoid line
136	point on inferior sternocleidomastoid line
137	upper point on posterior neck line
138	midpoint on posterior neck line
139	inferior point on posterior neck line
140	hair point just anterior to 107
147	posterior point of hair in profile

Appendix 9: Resemblance Ratings Paper, *J Forensic Sci*

Stephan, C. N. (2002). Do resemblance ratings measure the accuracy of facial approximations? *Journal of Forensic Sciences*, *47*(2), 239-243.