Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 159S (2006) S159–S163
Forensic Science International
www.elsevier.com/locate/forsciint

Assessing facial approximation accuracy: How do resemblance ratings of disparate faces compare to recognition tests?

C.N. Stephan [a,*], R.S. Arthur [b]

[a] Anatomy and Developmental Biology, The University of Queensland, Australia
[b] Anatomical Sciences, The University of Adelaide, Australia

Available online 14 March 2006

Abstract

In the past, the accuracy of facial approximations has been assessed by resemblance ratings (i.e., the comparison of a facial approximation directly to a target individual) and recognition tests (e.g., the comparison of a facial approximation to a photo array of faces including foils and a target individual). Recently, several research studies have indicated that recognition tests hold major strengths in contrast to resemblance ratings. However, resemblance ratings remain popularly employed and/or are given weighting when judging facial approximations, thus indicating that no consensus has been reached. This study aims to further investigate the matter by comparing the results of resemblance ratings and recognition tests for two facial approximations which clearly differed in their morphological appearance. One facial approximation was constructed by an experienced practitioner privy to the appearance of the target individual (practitioner had direct access to an antemortem frontal photograph during face construction), while the other facial approximation was constructed by a novice under blind conditions. Both facial approximations, whilst clearly morphologically different, were given similar resemblance scores even though recognition test results produced vastly different results. One facial approximation was correctly recognized almost without exception while the other was not correctly recognized above chance rates. These results suggest that resemblance ratings are insensitive measures of the accuracy of facial approximations and lend further weight to the use of recognition tests in facial approximation assessment.

Keywords: Forensic science; Craniofacial identification; Facial reconstruction; Facial reproduction; Face array; Quality assurance; Dexterity

1. Introduction

Methods of building faces from skulls, increasingly known as facial approximation [1–8], have long been regarded as useful techniques of last resort in aiding the identification of individuals in forensic cases. Over the past few years, empirical tests of methods have become more frequent and attempts at improving techniques and increasing accuracy have been made [9]. However, also crucial to the continued development of methods is the refinement of protocols used to assess the final constructed faces so that capabilities of methods and practitioners can be gauged and method areas conducive to improvement identified.

Presently, a number of methods have been employed to assess the accuracy of final faces produced by facial approximation. Resemblance rating is one such method where a measure of similarity is made between the facial approximation and a known target individual. This method appears to be the most frequently employed technique probably because it is quick and easy to conduct. Both quantitative and qualitative methods have been used in the resemblance rating approach. Qualitatively, facial approximations have been assessed by subjective statements regarding the likeness of the constructed face to the target individual [10–12]. For example, statements like ‘‘The reconstructed face bore an uncanny resemblance to the photograph [of the target individual]’’ are used to indicate the accuracy [11]. In quantitative methods, a metric scale is used to score the likeness of the facial approximation to the target individual. Commonly employed scales are from 1 to 5 [13–15] and 0 to10 [16,17].

An alternative to facial approximation accuracy assessment using resemblance rating methods is to assess the accuracy of facial approximations by conducting recognition tests where an assessor evaluates an array of face photos with reference to a facial approximation and then identifies a

* Corresponding author. Tel.: +61 7 3365 2958; fax: +61 7 3365 1299.
E-mail address: c.stephan@uq.edu.au (C.N. Stephan).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2006.02.026

face. Such tests appear to be less frequent than resemblance ratings, probably because they take more effort and time to assemble and conduct.

A variety of recognition tests exist since people who are unfamiliar or familiar with the target person may be studied and face arrays can be presented to these assessors either sequentially or simultaneously. The most common recognition tests employed to test facial approximations are ''unfamiliar tests'' where simultaneous face arrays are used [6,8,14,17,18]. Familiar recognition tests of facial approximation are infrequent since these usually require the participation of the next of kin of victims and thus may place participants in uncomfortable circumstances. However, there are examples where such tests could, or have been, conducted by using living individuals as the basis for facial approximation construction [8,11].

Research investigating the best method of facial approximation assessment has pointed to recognition tests as superior to resemblance ratings [16,17]. The evidence for this is that correctly recognized facial approximations are not given higher resemblance scores than incorrectly recognised facial approximations [16] and that facial approximations given high resemblance ratings may not be recognized above rates expected by chance [17]. The weakness of resemblance ratings is not perhaps surprising given that (i) resemblance ratings test the similarity of the faces to each other, not the ability for the facial approximation to be recognized (upon which facial approximation casework depends); (ii) resemblance rating tests fail to account for individuals who may bear closer resemblance to the facial approximation than the target; and (iii) anatomically similar faces are not be the only recognizable faces, see, e.g., recognition of caricatures, cartoons and pixelated images [19–21]. Support for recognition tests being superior to direct comparisons of a face to a target individual (as is done for resemblance ratings) is also evident from the eyewitness identification literature where direct comparisons are widely recognized to be biased [22–24].

In contrast to the above evidence, unpublished studies have recently been cited where the resemblance ratings of facial approximations to target individuals were found to be greater than the resemblance to non-target individuals [14]. It has also been reported that these facial approximations, which often gained a close resemblance to the target individual, were recognized above chance rates [14]. This suggests that there may be a large correlation between a high resemblance scores and correct recognition. However, these results may need to be regarded with some caution as the findings appear to be based on protocols that compare each facial approximation with only one preselected non-target individual [14] which may introduce bias. Irrespectively, the continued use of both resemblance ratings and recognition tests in the facial approximation literature suggests that further investigations are required. The aim of this study is to determine the usefulness of resemblance ratings for assessing facial approximation accuracy by comparing resemblance rating scores to recognition results of two facial approximations which clearly differ in terms of their morphological appearance (to each other and to the target face).

2. Materials and methods

Two facial approximations were constructed from casts of the same skull under conditions predisposing the production of faces with different appearances (Fig. 1). One facial approximation (FA1) was constructed by an experienced practitioner (CNS) who, after using combination methods of facial approximation to construct the face (see below), was given a photograph of the target individual and proceeded to alter the facial approximation to make it ''match'' the photograph. Thus, this facial approximation was not constructed under blind conditions, as the practitioner was privy to the appearance of the target individual. The second facial approximation was constructed by a novice (final year undergraduate science student – RSA) under blind conditions, using identical combination methods as for FA1. During face construction, Helmer's [25] soft tissue depths for 20- to 30-year-old females, rounded to the nearest 0.5 mm, were used at 10 landmarks in the median plane (opisthocranion; vertex; glabella; nasion; mid-nasal; subnasale; labrale superius; labrale inferius; pogonion; gnathion) and 10 bilateral landmarks (orbitale; mid-supraorbital; supra-second molar; infra-second molar; alare; zygomaxillare; ''mid-ramus''; zygion; ''border of mandible''; gonion). The general outline for three dimensional clay combination methods of facial approximation were followed according to Prag and Neave [11], but methods were supplemented with new guidelines described by Stephan [26,27]. Clay modelling was executed in 13.5 h for FA1, and 12 h for FA2.

The accuracy of each facial approximation was assessed under two conditions: (i) a resemblance rating test, and (ii) a recognition test using unfamiliar assessors and a simultaneous face array. Different assessors were used in each condition giving four sample groups. The second author (RSA) was responsible for conducting all facial approximation assessment trials, which included 166 participants in total. In the direct comparison tests, assessors were asked to give a resemblance rating from 0 to 5 with 0 indicating no resemblance and 5 indicating a high resemblance. In the recognition tests, assessors were asked to attempt to identify the target face from a simultaneous face array ($n = 10$) using the facial approximation. Participants were told that the target face may not be present in the array (even though it was) and, therefore, each participant had an average theoretical chance of 5% for selecting any face (50% chance of identifying a face multiplied by 10% chance of selecting any particular face). Fourteen adult assessors (males: $n = 4$; females: $n = 10$) were used for the direct comparison test of FA1, while 17 adult assessors (males: $n = 8$; females: $n = 9$) were used for the recognition tests. Eighty-four adult assessors (males: $n = 32$; females: $n = 52$) were used for the direct comparison test of FA2, while 51 adult assessors (males: $n = 10$; females: $n = 41$) were used for the recognition tests. All responses were analysed using Microsoft® Excel® 2000 and the JMP® 4.0 statistical packages.