PERGAMON

Archives of Oral Biology 47 (2002) 545–554

Archives of Oral Biology

www.elsevier.com/locate/archoralbio

Impact of landmark reliability on the planar Procrustes analysis of tooth shape

D.L. Robinson [a,b,*], P.G. Blackwell [b], E.C. Stillman [b], A.H. Brook [a]

[a] *Department of Child Dental Health, School of Clinical Dentistry, University of Sheffield, Claremont Crescent, Sheffield S10 2TA, UK*
[b] *Department of Probability and Statistics, Hicks Building, University of Sheffield, Hounsfield Road, Sheffield S3 7RH, UK*

Accepted 22 March 2002

Abstract

In a recent study, the ideas of Procrustes analysis were introduced to the study of tooth shape for teeth represented as configurations of 'landmarks' from digital images. This study aimed to establish how well the method could be expected to perform (in its standard form) when used on surfaces from a variety of tooth types and, in particular, how much impact inconsistencies in the positioning of landmarks would have on investigations of shape.

Using four different operators' images and landmarks from 10 different surfaces from each of 20 patients, the consequences of location inconsistency are evaluated by calculating its effect on the recorded variation in Procrustes fits, obtained for each set of multiple representations. The proportion of variation in shape attributable to actual differences between patients, rather than other sources of error, ranged from only 36 to 65% for the five buccal-surfaces considered and was no more than 30% for any of the five occlusal surfaces. Further examination of these results indicated that consistent orientation differences before imaging might be a particular source of error in obtaining any occlusal-landmark data, as might location ambiguities around the edges of the teeth. Orientation effects were also suggested for the buccal-surfaces of the molar teeth. In contrast, the relatively flatter buccal-surfaces of the incisors and canines produced the most reliable data.

Methods of analysis need to accommodate these problems if landmark data are to be used to describe variations in tooth shape. Different surfaces each present their own particular difficulties and so a variety of solutions may be required.

Keywords: Reliability; Reproducibility; Procrustes analysis; Tooth shape; Landmarks

1. Introduction

Detailed analysis of variation in tooth shape is important in studies of dental development but until recently has simply involved studying ratios of distances or angles between subjectively identified landmarks. A statistical framework, evolved from the ideas of Bookstein (1986) and Kendall (1984) and collated into the recent text by Dryden and Mardia (1998), now exists for the study of shape when objects are represented as configurations of 'landmarks'. In Robinson et al. (2001), we introduced the ideas of Procrustes analysis to the study of tooth shape. A key benefit is that the techniques allow work with the full geometry of teeth, which has previously been ignored.

Using digital images of buccal or occlusal surfaces from study models, the locations of subjectively identified landmarks may be recorded as configurations of two-dimensional (x, y) coordinates. Landmarks represent key points of correspondence, labelled and defined in the same way on each tooth of a given type, and are selected to be informative about characteristics of interest. Planar Procrustes analysis then allows estimates of mean shapes and visualizations of shape variability to be obtained and conventional, two-dimensional inferential techniques adapted to address hypotheses concerning shape.

However, for these methods to perform well, landmarks must be reliably located. Otherwise this can lead to problems later in an analysis. Inconsistencies between operators in the positioning of landmarks will result in inconsistent representations of shape. 'Real' differences will become diluted by increased residual variance and statistical

* Corresponding author. Fax: +44-114-2717843.
E-mail address: d.l.robonson@sheffield.ac.uk (D.L. Robinson).

0003-9969/02/$ – see front matter
PII: S0003-9969(02)00038-9

power will be reduced, lessening our ability to reject false hypotheses.

We have now sought to quantify the reliability of landmark positioning in terms of its impact on the planar Procrustes analysis of shape and therefore establish how useful a technique this is in its standard form for investigations of tooth shape. From buccal and occlusal views of a variety of tooth types, we establish the level of variation due to inter-operator inconsistency in obtaining landmarks, relative to actual variation in tooth shape. For the unfamiliar reader, the definitions and concepts of Procrustes analysis are briefly revisited.

2. Materials and methods

2.1. Sources of error

2.1.1. Data acquisition

To obtain landmark coordinates from dental study casts, the following of steps were required. Each model was secured on a platform adjustable in three planes beneath a mounted 32-bit digital camera (Kodak/Nikon DCS 410), used to acquire the images. Standardized definitions for the orientation of each tooth surface at the image-capture stage were used to aid correspondence. The criterion used was that each surface should be positioned parallel to the camera, so that the maximum surface area was visible. Once displayed on screen, the locations of landmarks were then recorded as (x, y) pixel coordinates in the two-dimensional plane of the image, using the Image ProPlus V 3.01 software package (Media Cybernetics). The x-axis ran parallel to the bottom horizontal edge of the image window and the perpendicular y-axis parallel to the left side of the window. Landmarks were positioned using a mouse-controlled icon, accurate to a single pixel and which could be repositioned at any time during the landmark identification, if required. Images were 1012 × 1524 pixels, displayed on a 17 in. monitor.

As an additional step, the axes of the image could also be calibrated by placing a rule in the plane of the tooth surface, before imaging, which is then used to specify the scale of the two axes. The accuracy of this calibration would be important to consider when assessing the impact of landmark inconsistencies on measures of size, but in an analysis of shape the scale of an object is irrelevant (see later).

Arnqvist and Martensson (1998) present a comprehensive description of the components of error that may be encountered when identifying and recording landmark locations on two-dimensional images of three-dimensional objects by the procedures described previously. They suggest that 'total error' may be partitioned into 'methodological' (from specimen presentation), 'instrumental' (optical or digital distortion) and 'personal' errors (subjective decisions). Whilst there may be methodological errors involved in obtaining impressions and forming study casts from patients, this part of the process will not be repeated for this study. The reproducibility of study models is not in question here and our assessments of reliability are only required to represent the subsequent components of error that are repeated. It is also expected that instrumental effects on the accuracy and precision of recorded data-points due to the lens distortion or other characteristics of the camera, digital transfer and representation on a computer screen are negligible (see manufacturer's service manuals). Consequently, we expect that the primary source of landmark errors will be in the operators' subjective positioning of points and/or due to subjective orientation of surfaces during image capture. By examining patterns of variation in the recorded shapes (see Section 3), it was hoped that we could identify which particular landmarks were affected in these ways and how, so that we might then consider ways in which reliability could be improved.

2.1.2. Operators

Four operators were available to take part in this study, allowing variations between individuals' data to be represented in the final results. A single operator may be highly successful in recording similar representations of a configuration on multiple occasions, but these representations must also agree with those of other operators using the technique, so that results may be validated or verified independently and so we may guard against systematic error. The operators were postgraduate students from the Department of Child Dental Health, each with over 2 years experience of using the imaging system and software. In the analysis that follows they are regarded as a typical sample from a larger population of trained operators, so that estimated reliability levels extend beyond just those who took part in this study.

2.2. Patients, surfaces and landmarks

In order to provide a comprehensive representation of the different surfaces found within the dentition and to encompass a variety of different tooth features, up to 20 of each of the surfaces listed later were imaged and represented as landmark configurations by each operator. Each tooth was required to be fully erupted, with no evidence of attrition and to have none of its surface obscured by crowding. The investigation made use of existing material currently under investigation in other projects with the Department of Child Dental Health for which ethics committee approval had been obtained. The 20 independent patients comprised 10 males and 10 females of a variety of different genetic and environmental backgrounds, such as may be of interest in some future study.

The surfaces considered are displayed, along with the definitions of landmarks used, in Fig. 1. The buccal-surfaces studied were: upper right central incisor, upper right first molar, lower left central incisor, lower left canine and lower left first molar. The occlusal surfaces used were: upper right central incisor, upper right first pre-molar, lower left canine,

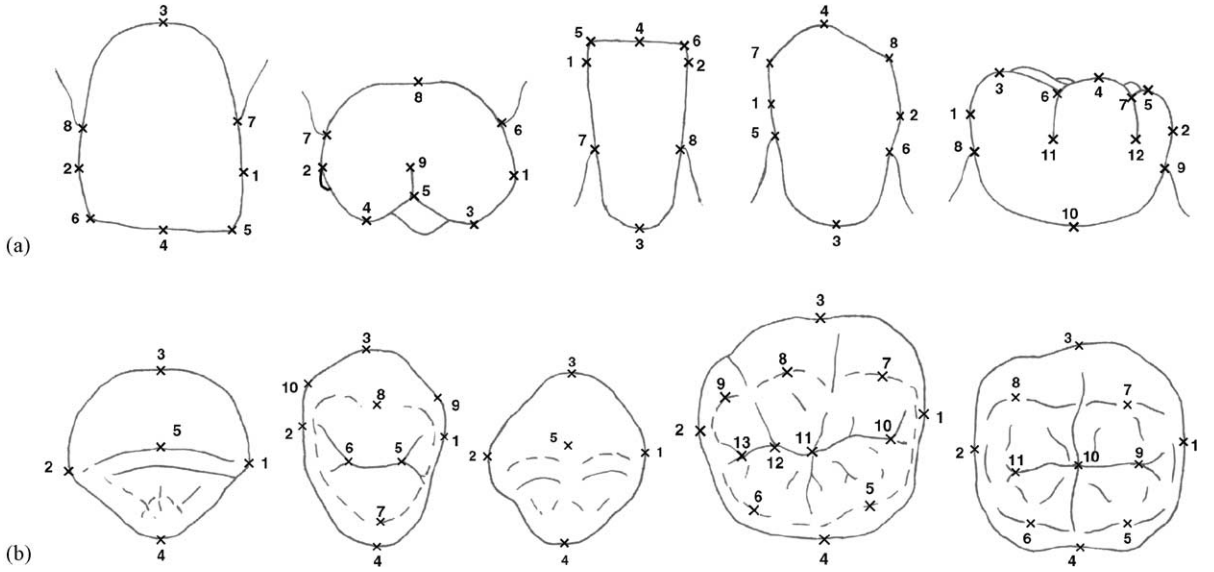

Fig. 1. (a) Buccal-surfaces and landmarks used. From left to right (upper right central incisor): (1) and (2) mesial and distal endpoints of MD; (3) and (4) gingival and incisal endpoints of the LACC; (5) and (6) corners of mesial and distal sides and incisal edge; (7) and (8) ends of mesial and distal papillae. Upper right first molar: (1) and (2) mesial and distal endpoint of MD; (3) and (4) mesial and distal labial cusp-tips; (5) occlusal limit of buccal groove; (6) and (7) ends of mesial and distal papillae; (8) half way between (6) and (7) along gingival margin; (9) start of buccal groove. Lower left central incisor: (1) and (2) mesial and distal endpoints of MD; (3) and (4) gingival and incisal endpoints of LACC; (5) and (6) corners of mesial and distal sides of tooth and incisal edge; (7) and (8) ends of mesial and distal papillae. Lower left canine: (1) and (2) mesial and distal endpoints of MD; (3) and (4) gingival and cusp-tip endpoints of LACC; (5) and (6) ends of mesial and distal papillae; (7) and (8) mesial and distal angles of cusp. Lower left first molar: (1) and (2) mesial and distal endpoints of MD; (3)–(5) mesial, central and distal cusp-tips; (6) and (7) occlusal limits of mesial and distal buccal grooves; (8) and (9) ends of mesial and distal papillae, (10) half way between 8 and 9 along gingival margin; (11) and (12) starts of mesial and distal buccal grooves. (b) Occlusal surfaces and landmarks used. From left to right (upper right central incisor): (1) and (2) mesial and distal endpoints of MD; (3) and (4) buccal and lingual endpoints of BL; (5) point of bisection of incisal edge by BL. Upper right first pre-molar: (1) and (2) mesial and distal endpoints of MD, (3) and (4) buccal and lingual endpoints of BL; (5) and (6) mesial and distal pits/fissure junctions; (7) and (8) lingual and labial cusp-tips; (9) and (10) mesial and distal endpoints of maximum labial cusp width. Lower left canine: (1) and (2) mesial and distal endpoints of MD; (3) and (4) buccal and lingual endpoints of BL; (5) cusp-tip. Lower left first molar: (1) and (2) mesial and distal endpoints of MD; (3) and (4) buccal and lingual endpoints of BL; (5) and (6) mesial and distal lingual cusp-tips; (7)–(9) mesial, central and distal labial cusp-tips; (10)–(12) and (13) outer mesial, inner mesial, central and distal pits. Lower left second molar: (1) and (2) mesial and distal endpoints of MD; (3) and (4) buccal and lingual endpoints of BL; (5) and (6) mesial and distal lingual cusp-tips; (7) and (8) mesial and distal labial cusp-tips; (9)–(11) mesial, central and distal pits.

lower left first molar and lower left second molar. The tooth outlines were created from the diagrams in Wheeler (1969).

The landmarks in Fig. 1 are primarily ‘anatomical’; points, assigned by dental clinicians, of homologous, meaningful, biological correspondence. They consist of features such as cusp-tips and fissure junctions in addition to positions corresponding to endpoints of commonly used clinical measurements such as the mesiodistal width (MD), buccolingual width (BL) and the long axis of the clinical crown (LACC). The MD was defined as the maximum diameter between the contact areas of a tooth. The BL was defined as the maximum buccolingual diameter in the occlusal view, approximately perpendicular to the MD. The LACC was to be placed to divide the buccal-surface roughly in half, in the occlusal–gingival direction. Points around the gingival margin were included for buccal views to provide some indication of the shape of teeth in these regions. In remaining regions with few characteristics of correspondence, points placed half way between other landmarks were used. These are known as ‘pseudo landmarks’.

2.3. Planar Procrustes analysis: concepts and definitions

2.3.1. Dissimilarity in shape

Shape is defined by Dryden and Mardia (1998) as ‘all the geometrical information that remains when location, scale and rotational effects are filtered out from an object’. Consequently, distances between landmark locations obtained from different images are meaningless unless the unwanted ‘registration’ differences are removed between configurations.

Therefore, in order to compare the shape of two corresponding landmark configurations and quantify how they differ, they are firstly each translated (to (0, 0)), so that their centres are coincident. They are then resized so that each centroid size (sum of squared distances of landmarks to the configuration centre) equals 1. One configuration is then rotated about (0, 0) until the sum of squared distances between corresponding landmarks is minimized. This minimized sum is known as the 'partial Procrustes' distance between the two shapes. It is then possible to reduce the sum of inter-landmark distances further, by resizing one shape to obtain the 'full Procrustes' distance between configurations. In practice, however, this makes very little difference (Dryden and Mardia, 1998). The procedure is illustrated in Fig. 2 for two buccal-surface landmark sets from images of two upper right central incisors.

2.3.2. Shape variation

An estimate of mean shape may be obtained from a sample of landmark configurations as the shape which has the least sum of Procrustes distances to each configuration in the sample. This can be found by a variety of methods (for example Kent (1994)). Fitting each sample configuration to the estimated mean shape gives its 'Procrustes fit', with Procrustes coordinates $(x_{ij}^{\mathrm{P}}, y_{ij}^{\mathrm{P}})$, where $j = 1, \ldots, k$ identifies the landmark number and i denotes the configuration. In each of our 10 samples, there will be up to 80 configurations as we have four operators' representations of each of $n \leq 20$ patients' teeth.

The structure of shape variability around the mean may be described by the calculation of principal components. In the current context, the technique summarizes the total variation in the Procrustes fits $(x_{ij}^{\mathrm{P}}, y_{ij}^{\mathrm{P}})$ in terms of uncorrelated, linear combinations of these coordinates. The first principal component obtained, PC_1, represents the largest amount of variation in shape, PC_2 the second largest and so on, until as many components are calculated as there are coordinates. Hence, for a surface represented by k landmarks there will be $2k$ components. Typically most of the sample variation is captured in the first few components.

'Component scores', values on each principal component for each configuration, may then be calculated. The score on PC_r, $r = 1, \ldots, 2k$ for configuration i, is given by $\sum_{j=1}^{k}(\gamma_{rjx}x_{ij}^{\mathrm{P}} + \gamma_{rjy}y_{ij}^{\mathrm{P}})$, where the pairs of weights γ_{rjx} and γ_{rjy} describe directions of variation in each Procrustes coordinate, about the mean shape. See, for example, Mardia et al. (1979) for a general description of principal component analysis.

Visualisation involves displaying hypothetical configurations at different extremities of shape variation, in the directions defined by the principal components. For PC_r, we can plot shapes with coordinates for each landmark j given by $(\bar{x}_j^{\mathrm{P}} + c\sqrt{\lambda_r}\gamma_{rjx},\ \ \bar{y}_j^{\mathrm{P}} + c\sqrt{\lambda_r}\gamma_{rjy})$, where $(\bar{x}_j^{\mathrm{P}},\ \ \bar{y}_j^{\mathrm{P}})$ are the coordinates of the Procrustes mean shape and λ_r is the variance captured by PC_r. Typically $c = -3, -2, \ldots, +2, +3$, so that configurations with scores between -3 and $+3$ S.D. either side of the mean shape may be obtained.

Figs. 3 and 4 (see Section 3) show these configurations for PC_1 and PC_2 obtained from the buccal and occlusal surfaces, respectively. For example, in Fig. 3 (top left), upper central incisor configurations with component scores of -3, -2, -1 and $+1$, $+2$, $+3$ S.D., either side of the mean shape indicate that the component describes variation in the shape of the gum and how this is related to the taper of the sides of the tooth.

2.4. Reliability assessment

The calculation of principal components is also relevant to the assessment of reliability. For each of the 10 surfaces, we measure the impact of operator inconsistency in landmark positioning by evaluating its effect on the variation in the Procrustes fits obtained from each set of multiple representations. By separating the variation in shape into uncorrelated variables using principal components analysis, we can quantify the reliability in each set of component scores data using well-established methods of assessment, identifying how much of recorded variation in shape is true (biological) variation and how much is variation in scores due to inconsistently located landmarks.

Consider the principal components extracted from the sample of upper central incisor (buccal-surface) configurations. On each component, there will be 4×20 component scores. Considering any one particular component, PC_r, the component score for patient (or case) i, $(i = 1, \ldots, 20)$, obtained by operator m, $(m = 1, \ldots, 4)$ may be written as $S_{r_{im}} = C_{r_i} + O_{r_m} + E_{r_{im}}$. Here C_{r_i} denotes the mean of many such (replicated) observations on the tooth, and so may be thought of as the cases' (patients') 'true score' for this component. Each C_{r_i} is estimated as the mean score for each patient, averaged over the operators. O_{r_m} represents the mth operator's systematic (constant) additive effect on the scores (mean 0) and $E_{r_{im}}$ the random (non-constant) part of the difference from the true score (mean 0) in this particular direction of variation. The terms are considered mutually independent with variances $\sigma_{C_r}^2$, $\sigma_{O_r}^2$ and $\sigma_{E_r}^2$, respectively and therefore

$$\mathrm{Var}(S_{r_{im}}) = \sigma_{C_r}^2 + \sigma_{O_r}^2 + \sigma_{E_r}^2$$

Estimates having been obtained of each component of variance, the interoperator coefficient of reliability (Fliess, 1986) may be calculated for this principal component. This expresses the relative magnitude of the variation between patient means (true scores) and is defined as

$$R(C)_r = \frac{\sigma_{C_r}^2}{\sigma_{C_r}^2 + \sigma_{O_r}^2 + \sigma_{E_r}^2}$$

i.e. actual variation between patients as a proportion of the total observed variation in the component scores.

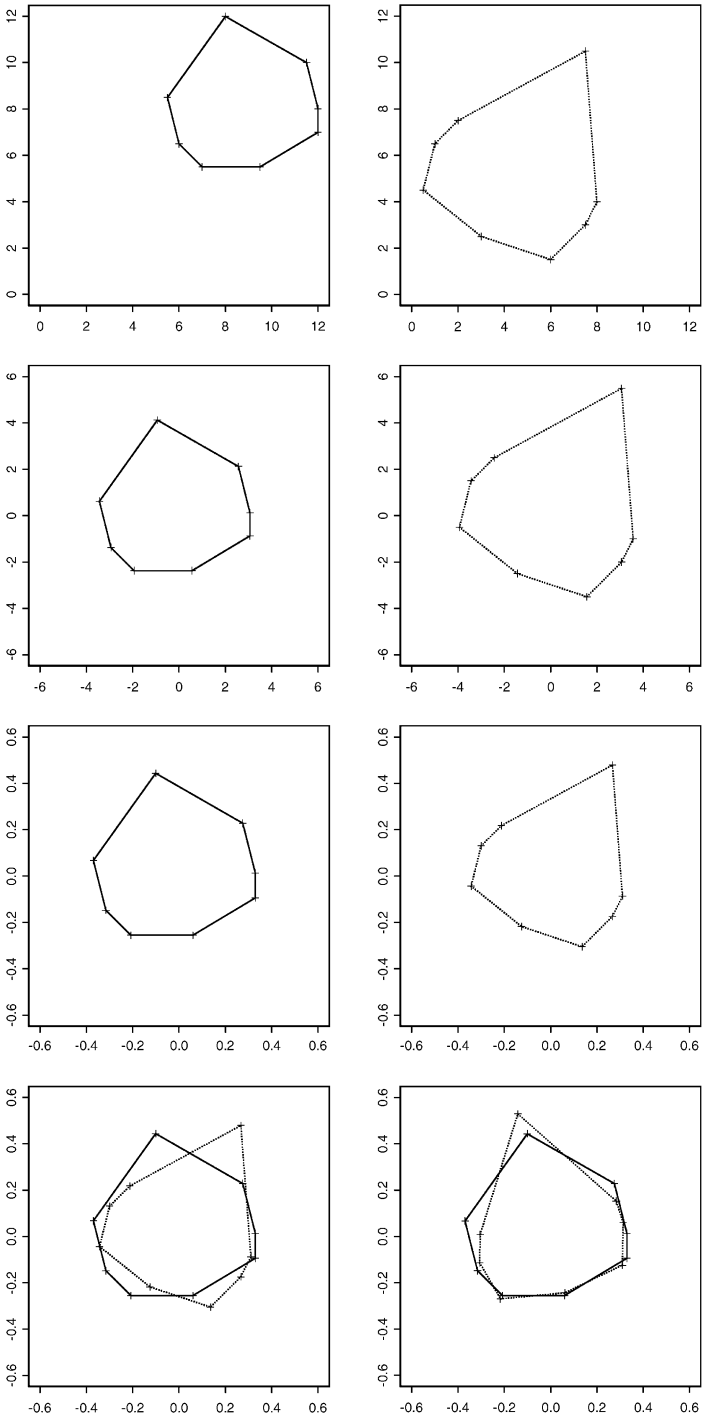

Fig. 2. Procrustes superimposition of two landmark configurations. First row: original recorded coordinates of each configuration on the image axes. Second row: centred and (third row) scaled version of each configuration. Final row: superimposed centred unit sized shapes rotated about (0, 0) to give the 'partial' Procrustes fit of configuration 2 (solid line) to configuration 1 (dotted line).

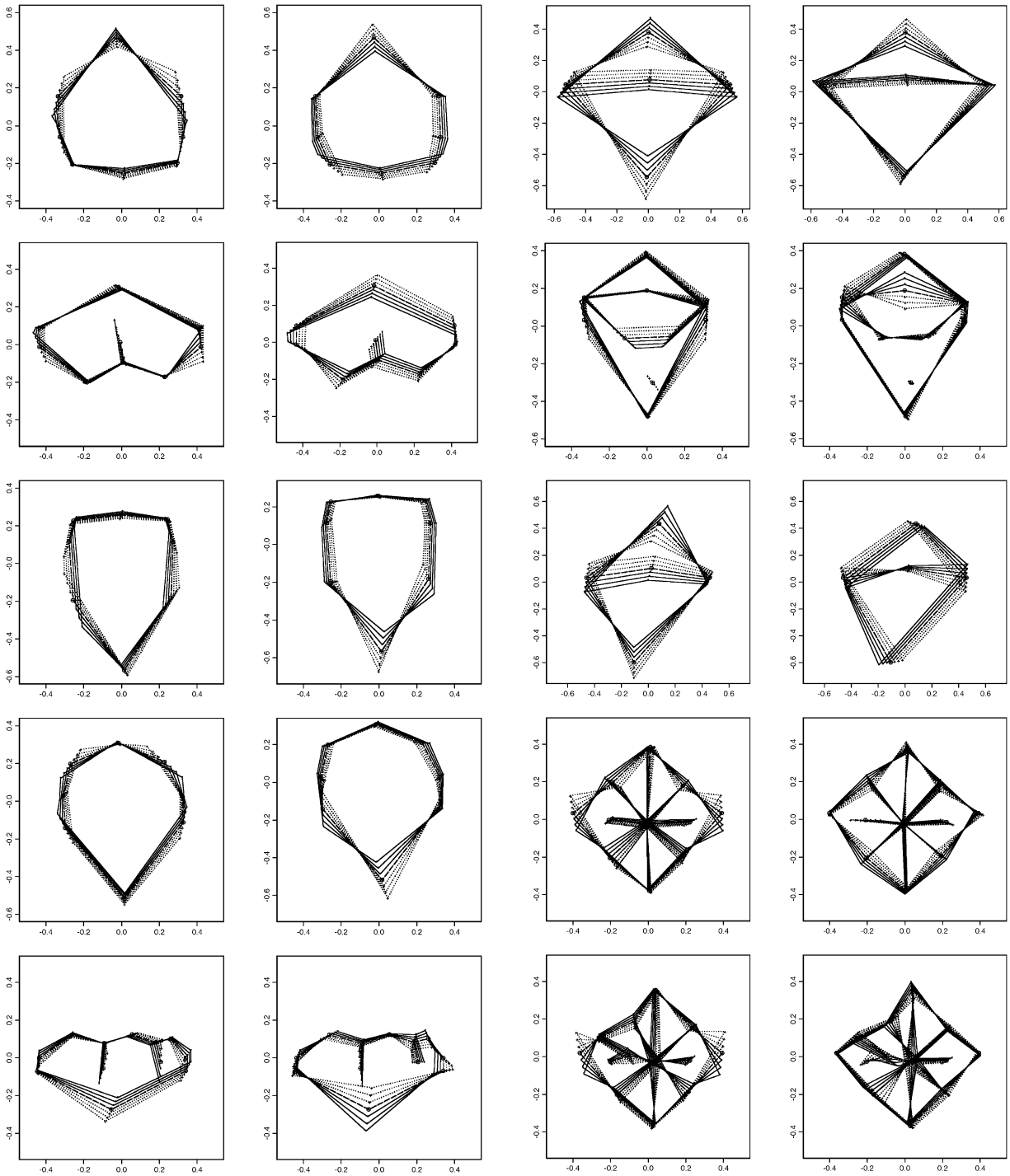

Fig. 3. First (left column) and second (right column) PC from buccal-surfaces. Lines represent configurations at −3, −2, −1 (dotted) and +1, +2, +3 (solid) S.D. either side of the mean shape (dashed), along each PC. Top row: upper right central incisor. Second row: upper right first molar. Middle row: lower left central incisor. Fourth row: lower left canine. Bottom row: lower left first molar. Rows 1 and 2: gingival landmarks at top, mesial landmarks on right. Rows 3–5: gingival landmarks at bottom, mesial landmarks on left.

Fig. 4. First (left column) and second (right column) *PC* from occlusal surfaces. Lines represent configurations at −3, −2, −1 (dotted) and +1, +2, +3 (solid) S.D. either side of the mean shape (dashed), along each PC. Top row: upper right central incisor. Second row: upper right first pre-molar. Middle row: lower left canine. Fourth row: lower left first molar. Bottom row: lower left second molar. All rows: labial landmarks at top, mesial landmarks on right.

If the variation due to inconsistency ($\sigma^2_{O_r} + \sigma^2_{E_r}$) is small relative to the variability between patients $\sigma^2_{C_r}$, reliability $R(C_r)$ in this direction of variation is high (with maximum value 1). Conversely if ($\sigma^2_{O_r} + \sigma^2_{E_r}$) is large relative to $\sigma^2_{C_r}$, reliability is low (with minimum value 0). Note that $1-R(C_r)$ gives the proportion of observed variance attributable to systematic and random errors.

Because the principal components are uncorrelated, an overall reliability figure may be produced for each surface. The total sample variation in shape in each set of repeated configurations is

$$\mathrm{Var}\,(\mathrm{PC}_1) + \mathrm{Var}\,(\mathrm{PC}_2) + \cdots + \mathrm{Var}\,(\mathrm{PC}_{2k}) = \sum_{r=1}^{2k} (\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}).$$

So an overall reliability score for all variation in shape may be calculated as

$$R(C_{\mathrm{All}}) = \frac{\sum_{r=1}^{2k} \sigma^2_{C_r}}{\sum_{r=1}^{2k} (\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r})}$$

Values of reliability have previously been characterized by qualitative benchmarks, such as 'slight' for values <0.2 (Donner and Eliasziw, 1987). While these are frequently quoted, they are only arbitrary classifications and so will be avoided here.

The proportion of variance attributable to systematic operator differences or attributable to random errors will also be quoted here. Individual component and overall measures are similarly defined as

$$R(O_r) \frac{\sigma^2_{O_r}}{\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}}, \qquad R(E_r) = \frac{\sigma^2_{E_r}}{\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}},$$

$$R(O_{\mathrm{All}}) = \frac{\sum_{r=1}^{2k} \sigma^2_{O_r}}{\sum_{r=1}^{2k} (\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r})},$$

$$R(E_{\mathrm{All}}) = \frac{\sum_{r=1}^{2k} \sigma^2_{E_r}}{\sum_{r=1}^{2k} (\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r})}$$

Note that for each set of surfaces, it should also be possible to compute a multivariate version of $R(C_r)$ directly from the Procrustes fits, based on the determinants of the between-patient, between-operator and within-patient covariance matrices. This would remove the need to calculate principal components, although here they provide an informative pictorial summary of the main patterns of variation, and we would expect equivalent results to those based on the calculations shown earlier.

Note also that using ANOVA to obtain estimates of the variance components sometimes produces negative values. In this event, it is usual to set these estimates to 0. The estimates for the remaining components may then be taken from the ANOVA results or computed again using an iterative procedure such as maximum-likelihood estimation, which re-estimates the remaining terms after setting the negative estimates to 0. The difference in estimates produced were found to be negligible (zero, to two decimal places) for the data considered here and made no difference to the reliability figures.

3. Results

For several reasons, it was not always possible to obtain four representations of each of the 20 tooth surfaces. For example, some patients had lost first permanent molars early. Cases where a complete set of different operators' data was unavailable were omitted from the study. The reduced sample sizes for these surfaces are indicated in Tables 1 and 2 in parentheses.

For each of the five buccal and five occlusal tooth surfaces, Tables 1 and 2 display the overall $R(C_{\mathrm{All}})$ and individual principal-component reliability figures $R(C_r)$. To save space, only components describing at least 5% of variance in shape are given. The remaining proportions of variance attributable to systematic operator differences, $R(O_r)$ and due to random error, $R(E_r)$, on each component and overall $R(O_{\mathrm{All}})$, $R(E_{\mathrm{All}})$ are also reported.

For both sets of surfaces, the proportion of variance on each principal component attributable to random error $R(E_r)$ generally increased with component number, whereas reliability scores $R(C_r)$ (between-patients) and the proportion of variance due to constant differences between operators $R(O_r)$ generally decreased. This was expected, as principal-component analysis selectively recovers non-random structure in its early components (Arnqvist and Martensson, 1998; Lougheed et al., 1991).

Figs. 3 and 4 display the first two principal components, representing the largest two patterns of variation in shape, for each of the five buccal and five occlusal surfaces. Where only one of $R(C_r)$ or $R(O_r)$, $r = 1, 2$ is high, the diagrams provide an indication of likely biological differences or systematic irregularities in operators' data, respectively.

3.1. Buccal-surfaces

The overall reliability figures, describing the proportion of variation in shape attributable to 'actual variation', ranged from 0.36 to 0.65. For the canine and two incisor surfaces, the variation observed was mostly between patients, $R(C_{\mathrm{All}})$ ranging from 0.52 to 0.65. The two molar surfaces produced lower results, with more than half of the observed variation in shape being attributable to systematic and random operator differences ($R(C_{\mathrm{All}}) = 0.36$ and 0.46).

$R(C_1)$ and $R(C_2)$ from the first two principal components were found to be fairly high for each of the upper and lower central incisors and the lower canine surfaces (0.52, 0.66 and 0.65, respectively for PC_1, 0.73,0.86 and 0.68 for PC_2), and so these components mainly describe actual variations in tooth shape (between patients). For both incisors, PC_1

Table 1
Buccal-surfaces: proportions of variance attributable to different sources on principal components describing >5% of variance[a]

	Upper central incisor(19)				Upper first molar (20)				Lower central incisor (19)				Lower canine (20)				Lower first molar (15)			
	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)
PC_1	0.52	0.24	0.35	41.70	0.07	0.70	0.23	28.30	0.66	0.15	0.19	40.30	0.65	0.07	0.28	30.55	0.67	0.23	0.10	32.96
PC_2	0.73	0.13	0.11	20.04	0.69	0.09	0.22	22.43	0.86	0.05	0.09	23.66	0.68	0.08	0.24	17.54	0.22	0.57	0.21	24.39
PC_3	0.35	0.16	0.52	12.99	0.48	0.16	0.36	15.04	0.66	0.01	0.33	14.23	0.63	0.17	0.20	16.42	0.49	0.14	0.37	8.41
PC_4	0.59	0.13	0.21	11.80	0.51	0.11	0.38	8.16	0.35	0.23	0.42	10.83	0.35	0.07	0.58	11.45	0.58	0.11	0.31	8.01
PC_5					0.29	0.03	0.68	6.39					0.35	0.37	0.28	8.55	0.58	0.00	0.42	7.18
PC_6													0.51	0.01	0.48	6.45				
All	0.52	0.19	0.29		0.36	0.28	0.36		0.65	0.11	0.25		0.55	0.11	0.34		0.46	0.25	0.29	

[a] Number of cases in parentheses.

Table 2
Occlusal surfaces: Proportions of variance attributable to different sources on principal components describing >5% of variance[a]

	Upper central incisor(19)				Upper first pre-molar(18)				Lower canine(20)				Lower first molar(15)				Lower second molar (20)			
	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)	$R(C_r)$	$R(O_r)$	$R(E_r)$	Variance (%)
PC_1	0.23	0.55	0.22	59.75	0.06	0.75	0.19	30.03	0.06	0.47	0.47	46.99	0.01	0.92	0.07	26.03	0.02	0.90	0.08	26.54
PC_2	0.65	0.18	0.17	22.04	0.43	0.14	0.43	15.79	0.25	0.36	0.39	23.71	0.18	0.34	0.48	15.61	0.28	0.36	0.36	13.80
PC_3	0.10	0.01	0.89	7.18	0.16	0.25	0.59	13.14	0.21	0.03	0.76	9.93	0.18	0.42	0.40	10.43	0.38	0.08	0.54	10.44
PC_4	0.13	0.00	0.87	6.34	0.17	0.02	0.81	8.12	0.07	0.00	0.93	8.16	0.33	0.05	0.62	7.85	0.26	0.14	0.60	8.93
PC_5					0.30	0.06	0.64	6.18	0.29	0.03	0.68	5.99	0.18	0.00	0.82	6.33	0.29	0.04	0.67	7.36
PC_6					0.12	0.00	0.88	5.28									0.37	0.21	0.42	6.93
All	0.30	0.38	0.32		0.18	0.34	0.48		0.14	0.32	0.54		0.15	0.38	0.48		0.20	0.37	0.44	

[a] No. of cases in parentheses.

in Fig. 3 describes the shape of the gum in relation to the taper of the sides of the teeth. If the papillae landmarks are located further towards the proximal edge, relative to the gingival endpoint of the LACC, the teeth are more tapered in shape. The first principal component for the lower canine describes how the cusp-tip is more pronounced when the distance between the mesial and distal corners is similar to the MD. The second component for each of the three surfaces mentioned so far indicates variation in the gingival endpoint of the LACC (and hence the shape of the gum) relative to the width. These teeth appear to be proportionally narrower when more of the crown is exposed. $R(C_1) = 0.67$ for the lower first molar, with PC_1 in Fig. 3 describing the association between the relative height of the tooth and the lengths of both the mesial and distal buccal grooves, the distal cusp being relatively larger (wider) when the teeth are flatter in shape. For the upper first molar, $R(C_2) = 0.69$ and PC_2 contrasts the relative widths of the mesial and distal cusps with the height of the visible surface. This would suggest that for 'shorter' teeth, the distal cusp will be smaller than the mesial, but the converse will be true when height is proportionally larger relative to its width.

For the upper first molar, $R(O_1) = 0.70$ and so PC_1 in Fig. 3 indicates that differences in the position of the MD relative to the length of the buccal groove are largely attributable to consistent operator differences. Individuals who consistently placed the mesiodistal diameter of the tooth more occlusally/gingivally, also tended to record that the buccal groove was longer/shorter. In addition, $R(O_2) = 0.57$ for the lower first molar, with PC_2 in Fig. 3 describing variation in the relative height of the tooth and the length of the buccal groove. One possible explanation for both these observations would be that operators consistently orient the surface differently in the occlusal–gingival direction, before identifying the affected positions on screen. An examination of the operators' images appeared to support this.

3.2. Occlusal surfaces

The overall reliability figures, $R(C_{All})$ in Table 2, are poor, all being less than the lowest calculated for the buccal-surfaces. In all but one dataset, more than 80% of occlusal variation in shape was found to be attributable to inconsistencies in the landmark representations ($R(C_{All}) < 0.20$). For the upper central incisor, this figure was little better ($R(C_{All}) = 0.30$).

The second principal component for the upper central incisor was the only direction of variation for any occlusal surface to offer reliable information on variation between patients ($R(C_2) = 0.65$). In Fig. 4, this contrasts the buccolingual dimension with the MD.

For all occusal surfaces, a large proportion of variance on the first principal component was attributable to operators' systematic errors. For the upper central incisor and lower canine PC_1 in Fig. 4 shows differences in the position of the incisal edge or cusp edge (and hence the mesiodistal diameter) along the BL. For the lower first molar and lower second molar, $R(O_2) > 90\%$. Each of the first components in Fig. 4 contrasts the positions of all four cusp-tips, relative to the endpoints of the MD. The cusp-tips are located more lingually when the MD is placed more towards the outer buccal-surface. For the upper first pre-molar, systematic operator differences existed along PC_1 in the position of the MD and the position of the fissure, in the buccolingual direction. When operators placed the mesiodistal diameter nearer towards the buccal cusp, they also located the fissure back towards the lingual cusp. This and all the observations from the first principal components would most likely be due to individuals' consistent orientation differences in the imaging of this surface in the buccolingual direction, or where operators place the MD having acquired the image. For the upper first pre-molar, one would also expect that the position of the labial cusp-tip, relative to the buccal endpoint of the BL, would also be affected by orientation differences. This is represented in the second principal component for this tooth in Fig. 4, as actual and non-consistent operator variation ($R(C_2) = 0.43$, $R(E_2) = 0.43$), along with changes in the distal position of the maximum cusp width. Note that here, and on the other second components not discussed, the variations in scores are not solely attributable to just one source and so interpretation is not as straightforward.

4. Discussion

The tables indicate how reliable any single operator would be in representing the shape of a tooth surface if he/she were to carry out a future Procrustes investigation of shape, using these landmark sets. Reliability was generally higher for the buccal-surfaces than occlusal, with the canine and two incisor teeth producing the most encouraging results. In Robinson et al. (2001), a significant difference in mean buccal-surface shape was found when comparing the central incisors of a group of patients with hypodontia and a corresponding control group. However, while some statistical power remains, the overall figures for these surfaces still suggest the potential for problems, particularly where smaller differences in shape are to be investigated.

For the occlusal surfaces, overall reliability was poor (<0.3), the largest principal components suggesting that consistent orientation differences before imaging in the buccolingual direction could be a particular factor. Orientation effects were also suggested for the buccal-surfaces of the molar teeth, in the occlusal–gingival direction. One would expect that small changes in the orientation of occlusal surfaces would have a greater impact on the recorded landmark configurations than in the buccal view, as these surfaces are considerably more three-dimensional. Looking down on to an occlusal surface, slight movements cause features such as cusp-tips and incisal edges, which are nearer to the camera, to move around more, than other points around the gum and edges of the tooth, which are further away, such as the

endpoints of the BL and MDs. In contrast, the relatively flatter, more two-dimensional buccal-surfaces would not suffer from this problem to such an extent. An examination of the images for the occlusal surfaces did appear to confirm greater variation in orientation than others, which would clearly influence their representation as configurations of landmarks. Consequently, further standardization for positioning these surfaces at the imaging stage might reduce this source of variation.

However, whilst 'orientation' is clearly a problem for comparing configurations of landmarks, interpoint distances such as the MD or BL may still produce reliable measurements. Work is currently in progress to quantify the reproducibility of linear measurements based on the same sample of teeth and using the imaging system described here.

In each of the 10 datasets, the principal components were calculated from all the variation in shape represented in the Procrustes fits and consequently each set of component scores contained some variance between patients and some due to operator inconsistencies. The previous descriptions suggest which patterns of variation contained large proportions of error, but do not directly indicate the actual areas of systematic and non-systematic operator inconsistency. A further investigation of operator inconsistencies needs to follow if landmark techniques are to be used or adapted further to investigate actual variations in tooth shape. An examination of the 'within-case (patient)' and 'between-operator' covariance structures in the Procrustes fits should help identify further sources and regions of unwanted variation. Improved definitions for the location of landmarks might then help reduce positioning inconsistencies in some of these points. Operators also reported more difficulty in identifying certain landmarks than others on screen, such as (for the central incisor) those at the corners of the incisal edge and the MD. This issue also has implications in an analysis where each landmark carries equal 'importance'.

In addition to recording inconsistencies, other 'nuisance' variation results from differences between patients in the position of the gingival margin, regardless of the reproducibility of these points. Teeth may have the same shape but differences when represented as landmark configurations due to differences in the position of the gum around each tooth. The reliability figures here have no way of taking this into account. Any such differences in these regions will still be regarded as actual variation in shape, but this would be of no interest.

These landmarks are still useful, however, because they provide the best indication we have of the cementum–enamel junction and the relative dimensions of the teeth in these areas. Similarly, points that do not match exactly, but lie (say) around the outline of a tooth, still describe differences in shape perpendicular to these directions.

Methods of analysis need to accommodate the problems discussed here, using only information from the visible edges and surfaces of a tooth. Different surfaces each present their own particular difficulties and so a variety of solutions may be required if we are to develop a coherent methodology for the analysis of tooth shape. Current work is investigating the use of 'semi-landmarks' (Bookstein, 1996), an extension of the standard Procrustes methodology which recognizes that landmarks may be known to lie along particular lines or curves, such as around the outline or sides of a tooth (papillae landmarks), but are difficult to locate precisely. The technique allows these landmarks to move in the specified directions to overcome the lack of precise correspondence. It is hoped that the information gained from the datasets here will help in our development of these ideas and aid other future investigations of tooth shape.

Acknowledgements

We would like to thank the four postgraduate students who provided data for this investigation: Mohammad Al-Sharood, Mayada Karmo, Khaled Khalaf and Anwar Shah.

References

Arnqvist, G., Martensson, T., 1998. Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. In: Kingenburg, C.P., Bookstein, F.L. (Eds.), Acta Zoologica, Vol. 44 (1/2). Academaiae Scientarum Hungaricae, pp. 73–96.

Bookstein, F.L., 1986. Size and shape spaces for landmark data in two dimensions (with discussion). Stat. Sci. 1, 181–242.

Bookstein, F.L., 1996. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image Anal. 1, 225–243.

Donner, A., Eliasziw, M., 1987. Sample size requirements for reliability studies. Stat. Med. 6, 441–448.

Dryden, I.L., Mardia, K.V., 1998. Statistical Shape Analysis. Wiley, New York.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. Wiley, London.

Kendall, D.G., 1984. Shape manifolds, Procrustean metrics and complex projective spaces. Bull. Lond. Math. Soc. 16, 81–121.

Kent, J.T., 1994. The complex Bingham distribution and shape analysis. J. R. Stat. Soc., Ser. B 56, 285–299.

Lougheed, S.C., Arnold, T.W., Bailey, R.C., 1991. Measurement error of external skeletal variables in birds and its effect on principal components. Auk, 108, 432–436.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London.

Robinson, D.L., Blackwell, P.G., Stillman, E.C., Brook, A.H., 2001. Planar Procrustes analysis of tooth shape. Arch. Oral. Biol. 44, 191–199.

Wheeler, R.C., 1969. An Atlas of Tooth Form, 4th Edition. Saunders, Philadelphia.