

A Probability Distribution of Shape for the Dental Maxillary Arch Using Digital Images

Omar M. Rijal, *Member, IEEE*, Norli A. Abdullah, Zakiah M. Isa, Norliza M. Noor, *Senior Member, IEEE*, Omar F. Tawfiq

Abstract— Selected landmarks from each of 47 maxillary dental casts were used to define a Cartesian-coordinate system from which the positions of selected teeth were determined on standardized digital images. The position of the i -th tooth was defined by a line of length (l_i) joining the tooth to the origin, and the angle (θ_i) of this line to the horizontal Cartesian axis. Four teeth, the central incisor, lateral incisor, canine and first molar were selected and their position were collectively used to represent the shape of the dental arch. A pilot study using clustering and principal component analysis strongly suggest the existence of 3 groups of arch shape. In this study, the homogeneity of the 3 groups was further investigated and confirmed by the Dunn and Davies-Bouldein validity indices. This is followed by an investigation of the probability distribution of these 3 groups. The main result of this study suggests 3 groups of multivariate (MV) normal distribution. The MV normal probability distribution of these groups may be used in further studies to investigate the issues of variation of arch shape, which is fundamental to the practice of prosthodontics and orthodontics.

I. INTRODUCTION

Representation, clustering and evaluating changes of dental arch shape have long been of interest to the dentist. Proposing a representation of dental arch shape is a first step in finding the ‘ideal’ arch [1, 2, 3], comparing and clustering arches according to ethnicity, gender, size and shape [4, 5, 6], and evaluating arch changes before, during and after the dental treatment [7, 8, 9].

In a pilot study [10], two landmarks; the hamular notches and incisive papilla were used to define the Cartesian axes as well as the origin for the i -th maxillary dental cast from

selected digital images (Fig. 1). The Cartesian axes enable the position of any tooth, to be defined by a line of length (l_i) joining the tooth to the origin, and the angle (θ_i) of this line to the horizontal Cartesian axis. Test of symmetry of the maxillary arch in the pilot study essentially allows the position of four teeth which are the central incisor, lateral incisor, canine and first molar to represent the arch shape. Therefore the maxillary arch shape of 47 patients will be denoted by $\underline{v}_i = (\theta_{i1}, l_{i1}, \theta_{i2}, l_{i2}, \theta_{i3}, l_{i3}, \theta_{i4}, l_{i4}), i = 1, \dots, 47$.

Clusters of the vectors \underline{v}_i , were obtained from the agglomerative hierarchical clustering method (Fig. 2). Transformation of \underline{v}_i to its principal components show the existence of three clusters, with means $\bar{\underline{v}}_k = (\bar{\theta}_1^k, \bar{l}_1^k, \bar{\theta}_2^k, \bar{l}_2^k, \bar{\theta}_3^k, \bar{l}_3^k, \bar{\theta}_4^k, \bar{l}_4^k), k = 1, 2, 3$.

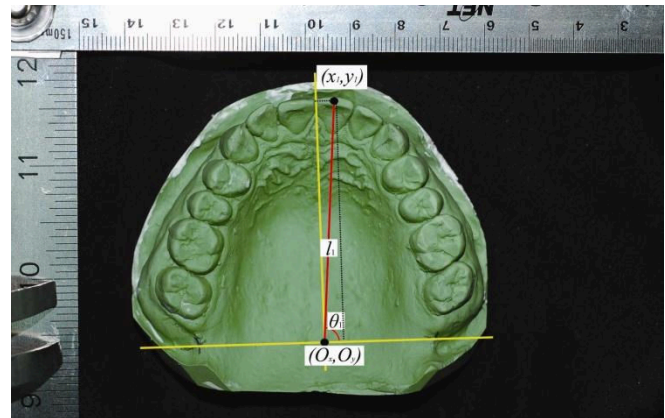


Figure 1. Computation of angle and distance of the tooth from the geometrical Cartesian origin made on digital images of casts positioned and calibrated in a standardized manner using two metal rulers positioned on a plane parallel to the occlusal plane.

In this study, distinction of the three clusters was verified using Dunns and Davies Bouldin validity indices. Distinct groups or homogenous group with respect to ethnicity, gender or shape was then investigated. Even if there is homogeneity of clusters, it is very unlikely that an individual \underline{v}_i will be equal to $\bar{\underline{v}}_k$. Hence, variation of \underline{v}_i with respect to $\bar{\underline{v}}_k$ is a more relevant issue to look at. The knowledge of $\bar{\underline{v}}_k$ and the distribution of \underline{v}_i is useful in the design of impression trays and may guide the prosthodontic

*Research supported by the Institute of Research Management and Monitoring (IPPP), University of Malaya and Ministry of Higher Education (MOHE) under SLAI program.

O. M. Rijal is with the Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia (e-mail: omarrija@um.edu.my).

N. A. Abdullah is with the Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia (corresponding author to provide phone: +6013-394-2514 (mobile); fax: +603-7967-4143; e-mail: norlienih@yahoo.com).

Z. M. Isa is with Department of Prosthetic Dentistry, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia (e-mail: zakiah@um.edu.my).

N. M. Noor is with the Department of Electrical Engineering, UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, Malaysia (e-mail: norliza@ic.utm.my).

O. T. Farouq is with Department of Prosthetic Dentistry, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia (e-mail: dromar_s@yahoo.com).

to construct the occlusion wax rim prior to making an appropriate denture when treating an edentulous patient. An objective way of investigating this variation is by determining the probability distribution of shape for each group.

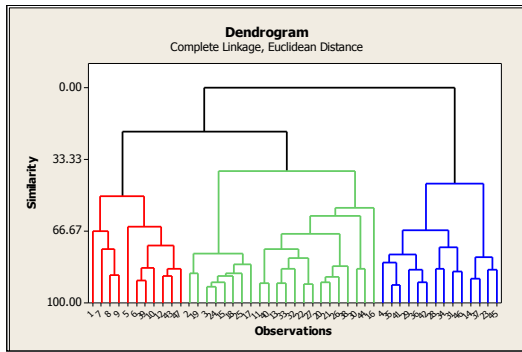


Figure 2. Clusters of the dental arch using the complete linkage.

II. MATERIALS AND METHODS

A. Data used in the study

Forty seven dentate adults aged between 19 and 32 years with well aligned maxillary anterior teeth and minimal attrition participated in this study. Subjects were excluded if they had a history of orthodontic treatment, anterior restoration or a fixed dental prosthesis in the maxilla or mandible. Impressions of dental and oral structures of the maxillary arches of the subjects were made with irreversible hydrocolloid (Duplast fast set alginate impression material; Dentsply Dental Co Ltd, Tianjin, China). Impressions were then cast using type III dental stone (Moldano; Heraeus Kulzer GmbH, Hanau, Germany).

B. Dunn and Davies Bouldein (DB) Validity Indices

To validate the separation and compactness of each group, Dunn's and DB validity indices were used [11]. Let c be the number of clusters. The Dunn's validity index, D is defined as:

$$D = \min_{1 \leq k \leq c} \left(\min_{1 \leq m \leq c} \left[\frac{\delta(\underline{v}_k, \underline{v}_m)}{\max_{1 \leq k \leq c} \{\Delta(\underline{v}_k)\}} \right] \right) \quad (1)$$

where $\Delta(\underline{v}_k) = \max_{v_i, v_j \in \underline{v}_k} \{d(v_i, v_j)\}$,

$\delta(\underline{v}_k, \underline{v}_m) = \max_{v_i \in \underline{v}_k, v_j \in \underline{v}_m} \{d(v_i, v_j)\}$,

$d(v_i, v_j)$ is the Euclidean distance of v_i and v_j and $i \neq j$.

In order to have clusters that are compact and well separated, the minimum intra-cluster distance (compactness) and maximum inter-cluster distance are required. Thus, from (1), the number of clusters that maximizes D suggests the optimal value of c .

Similar to the Dunn's index, the Davies-Bouldin index looks for sets of clusters that are compact and well

separated. It uses the idea based on the intra and inter-cluster distance as follows:

$$DB = \frac{1}{c} \sum_{k=1}^c \left(\max_{k \neq m} \left[\frac{\Delta(\underline{v}_k) + \Delta(\underline{v}_m)}{\delta(\underline{v}_k, \underline{v}_m)} \right] \right)$$

The number of clusters that minimizes DB suggests an optimal value of c , which is not necessarily the same number of clusters as suggested by the Dunn's index.

C. Test of Multivariate (MV) Normality Distribution

Initially, each component of \underline{v} was investigated for univariate normality using the Kolmogorov-Smirnov (KS) test [12, 13]. Once all components exhibit univariate normality, the Mardia MV test for skewness and kurtosis were carried out [14]. Let the null hypothesis be that the sample follows MV normality. The test statistics of skewness and kurtosis for a MV normal distribution are

$$A = \frac{nb_{1,p}}{6} \text{ and}$$

$$B = \frac{b_{2,p} - p(p+2)}{\sqrt{8p(p+2)/n}},$$

where $b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [(\underline{v}_i - \underline{v})' S^{-1} (\underline{v}_j - \underline{v})]^3$,

$b_{2,p} = \frac{1}{n} \sum_{i=1}^n [(\underline{v}_i - \underline{v})' S^{-1} (\underline{v}_i - \underline{v})]^2$, S is the sample covariance

matrix and p is the dimension of \underline{v} . It can be shown that A is asymptotically tends to a chi-squared probability distribution with $\frac{p(p+1)(p+2)}{6}$ degrees of freedom and

B is a standard normal random variable.

D. Comparing means from two multivariate normal populations using Hotelling T^2 test

The investigation of group separation was carried out with the Hotelling T^2 test [15]. Let the null hypothesis be $\mu_1 = \mu_2$. The Hotelling two sample statistics is

$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\underline{v}_1 - \underline{v}_2)' \mathbf{S}^{-1} (\underline{v}_1 - \underline{v}_2)$ with critical value

$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$, where \mathbf{S} is the covariance matrix

when the null hypothesis is true.

E. Applications of the above methods.

The Dunn's and DB validity indices will be used to investigate compactness and separation of clusters. Individual cluster that are sufficiently separated will be tested for normality and the existence of the normal distribution allows a formal method of testing group uniqueness.

III. RESULTS AND DISCUSSION

Agglomerative hierarchical clustering and principal component analysis using feature vector \underline{v}_i suggest three clusters [10]. However, principal components are not scale invariant. Therefore, Dunns and DB validity indices were used to validate the clusters and the result is given in Table I. The Dunn's index appears to be optimum at three clusters while DB at six clusters. However, the clustering method shows three clusters of 33% similarity level to each other while the six clusters give 60% similarity (Fig. 2). This suggests the preference of three clusters.

TABLE I. DUNN AND DB VALIDITY INDICES

Number of clusters, c	Dunn Index	DB Index
2 Clusters	1.2561	0.6753
3 Clusters	1.2977 (max)	0.4639
4 Clusters	1.1064	0.4274
5 Clusters	1.0365	0.2704
6 Clusters	1.1163	0.2108 (min)

TABLE II. ETHNIC AND GENDER HOMOGENEITY FOR EACH CLUSTERS.

Cluster	Ethnic	Percentage of ethnicity	Gender	Percentage of gender
\underline{v}_1	Malay ($n=6$)	54.54%	Male ($n=2$)	18.18%
	Chinese ($n=5$)	45.45%	Female ($n=9$)	81.81%
\underline{v}_2	Malay($n=18$)	81.81%	Male ($n=5$)	22.72%
	Chinese ($n=4$)	18.18%	Female($n=17$)	77.27%
\underline{v}_3	Malay ($n=10$)	71.42%	Male ($n=7$)	50%
	Chinese ($n=4$)	28.57%	Female ($n=7$)	50%

TABLE III. KOLMOGOROV-SMIRNOV TEST FOR UNIVARIATE NORMALITY. CRITICAL VALUES OBTAINED FROM [13] AT $\alpha = 0.05$.

Variables	Cluster 1 CV=0.3912	Cluster 2 CV=0.2809	Cluster 3 CV=0.3489
l_1	0.1800	0.1392	0.1787
θ_1	0.0946	0.1070	0.1425
l_2	0.1457	0.0926	0.2213
θ_2	0.2298	0.1178	0.1534
l_3	0.1642	0.1360	0.1927
θ_3	0.1927	0.1663	0.1400
l_4	0.2293	0.1259	0.1709
θ_4	0.0980	0.1380	0.1683

TABLE IV. MARDIA TEST OF SKEWNESS AND KURTOSIS CONFIRM THE MV NORMALITY DISTRIBUTION OF \underline{v}_k

Cluster	Mardia skewness (CV = 146.5674)	Mardia kurtosis (CV = 1.6449)
\underline{v}_1	69.8993	-3.1757
\underline{v}_2	96.4387	-2.0310
\underline{v}_3	89.8425	-2.4242

TABLE V. HOTELLING T^2 TEST FOR COMPARING TWO MULTIVARIATE MEAN.

Mean vector	T^2 statistics	F critical value
$\underline{v}_1 - \underline{v}_2$	75.0389	7.2618
$\underline{v}_1 - \underline{v}_3$	106.5962	8.8830
$\underline{v}_2 - \underline{v}_3$	107.5205	10.5000

Table II gives the percentage of ethnicity and gender for each cluster. Homogeneity in ethnicity and gender do not appear in each cluster, further suggesting that useful information may be derived by well defined groups of shape.

Confirmation of three clusters is done by the Hotelling T^2 test. However, before this is done, normality of the clusters must be proven. The KS test as illustrated in Table III shows univariate normality for all variables, which is the strong suggestion for MV normality. The Mardia test of skewness and kurtosis confirm the MV normality of \underline{v} (Table IV). Henceforth, the Hotelling T^2 test investigates pair wise differences of the mean vectors. Table V shows that all pairs of distribution are distinct.

It must be emphasized that MV normality of a random vector \underline{v} implies that each of its element is a continuous variable. Although θ_i is a circular measure, [16, 17] have shown that it can be approximated by linear measures.

This study has shown that three categories of shape can be identified by three MV normal distributions, namely,

$$C_j \sim MVN (\underline{v}_j, S_j); \quad j = 1, 2, 3$$

where

$$\underline{v}_1 = (83.16, 57.30, 74.85, 54.91, 67.27, 51.49, 36.82, 36.43),$$

$$\underline{v}_2 = (85.55, 53.55, 77.07, 51.61, 69.15, 48.35, 36.11, 33.91),$$

$$\underline{v}_3 = (87.30, 59.51, 79.43, 57.13, 72.21, 53.30, 43.76, 36.97),$$

$$S_1 = \begin{bmatrix} 2.01 & 0.33 & 2.05 & 0.39 & 1.37 & 0.59 & 1.37 & -0.15 \\ . & 2.92 & 0.38 & 2.21 & 0.58 & 2.05 & 0.59 & 0.52 \\ . & . & 2.24 & 0.47 & 1.78 & 0.57 & 2.21 & 0.19 \\ . & . & . & 2.04 & 0.88 & 1.88 & 1.27 & 1.20 \\ . & . & . & . & 2.57 & 0.87 & 3.89 & 1.85 \\ . & . & . & . & . & 1.95 & 1.10 & 1.12 \\ . & . & . & . & . & . & 7.98 & 3.40 \\ . & . & . & . & . & . & . & 4.47 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 1.78 & 1.04 & 1.95 & 0.78 & 2.06 & 0.58 & 1.83 & -0.21 \\ . & 4.62 & 1.53 & 3.84 & 1.70 & 3.75 & 3.37 & 1.26 \\ . & . & 2.34 & 1.27 & 2.53 & 1.20 & 2.90 & 0.15 \\ . & . & . & 3.73 & 1.65 & 3.78 & 4.00 & 1.84 \\ . & . & . & . & 3.01 & 1.63 & 3.91 & 0.51 \\ . & . & . & . & . & 4.12 & 4.39 & 2.23 \\ . & . & . & . & . & . & 9.43 & 2.98 \\ . & . & . & . & . & . & . & 3.33 \end{bmatrix} \text{ and}$$

$$S_3 = \begin{bmatrix} 3.70 & -2.99 & 3.50 & -3.10 & 2.79 & -2.91 & 1.52 & -1.72 \\ . & 4.83 & -2.54 & 3.97 & -2.04 & 3.85 & 0.57 & 1.68 \\ . & . & 3.54 & -2.74 & 2.96 & -2.51 & 2.08 & -1.64 \\ . & . & . & 4.06 & -2.20 & 3.89 & -0.23 & 1.99 \\ . & . & . & . & 2.64 & -2.01 & 1.95 & -1.37 \\ . & . & . & . & . & 4.11 & 0.04 & 2.03 \\ . & . & . & . & . & . & 6.33 & -1.52 \\ . & . & . & . & . & . & . & 2.73 \end{bmatrix}$$

The mean vector \bar{v}_k gives the knowledge of 3 distinct arch shapes. This information will help in the design of the impression trays that may generally accommodate a specific population and consequently facilitate the selection of impression trays for an inexperienced dentist.

In addition, knowledge of feature vector \bar{v}_j and the 3 shapes of dental arch may help the prosthodontists in treating edentulous patients. A formal method of developing a wax rim to replace the shrunken alveolar bone may be done using the 3 'average' arch shape. Determination of the teeth position may be then followed by using \bar{v}_1, \bar{v}_2 or \bar{v}_3 accordingly.

Further, the information of covariance S for each cluster gives the knowledge of variation of arch shape that is useful for diagnosis and treatment planning in orthodontics [1, 2, 3].

The limitation of the current study is mainly in the sample size. As such, to justify the clustering result, the Dunn's and DB validity indices were applied. As for the test of normality, several types of tests were applied and all of them confirmed normality. Existing efforts are now carried out to expand the sample size. Given the existing sample size, the results cannot be generalized to the Malaysian population.

ACKNOWLEDGEMENT

This project is supported by the Institute of Research Management and Monitoring (IPPP), University of Malaya and Ministry of Higher Education (MOHE) under SLAI program.

REFERENCES

- [1] M. A. MacConnail, E. A. Scher, "The ideal arch form of the human dental arcade with some prosthetic application," *Dent. Rec.*, vol. 69, pp. 285-302, 1949.
- [2] M. Raberin, B. Lai, J. L. Martom, F. Brunner, "Dimensions and form of dental arches in subjects with normal occlusions" *Am J Orthod Dentofac Orthop*, vol. 104, pp. 67-72, 1993.
- [3] P. D. Sampson, "Statistical analysis of arch shape with conic sections" *Biometrics*, vol. 39 (2), pp. 411-423, 1983.
- [4] B. G. Burris, E. F. Harris "Maxillary Arch Size and Shape in American Blacks and Whites" *Angle Orthod*, vol.70, pp. 297-302, 2000
- [5] K. Nojima, R. P. McLaughlin, Y. Isshiki, P. M. Sinclair. "A Comparative Study of Caucasian and Japanese Mandibular Clinical Arch Forms" *Angle Orthod*, vol.71, pp.195-200, 2001.
- [6] S. J. Lee, S. Lee, J. Lim, H. J. Park, and T. T. Wheelere "Method to classify dental arch forms" *Am J Orthod Dentofacial Orthop* vol. 140, pp. 87-96, 2011.
- [7] S. Mutinelli, M. Cozzani, M. Manfredi, M. Bee and G. Siciliani "Dental arch changes following rapid maxillary expansion" *European Journal of Orthodontics*, vol. 30, pp. 469-476, 2008
- [8] BeGole E. A. and Lyew R. C., "A new method for analyzing change arch form" *Am J Orthod Dentofacial Orthop* vol. 113, pp. 394-401, 1998
- [9] J. Baluta and C. L. B. Lavelle, "An analysis of dental arch form" *European Journal of Orthodontics*, vol. 9, pp. 165-171, 1987.
- [10] O. M. Rijal, N. A. Abdullah, Z. M. Isa, F. A. Davaei, N. M. Noor, O. F. Tawfiq "A Novel Shape Representation of the Dental Arch and its Applications in Some Dentistry Problems," In *Proc. 33rd Annual International Conference of the IEEE EMBS*, Boston, 2011, pp. 5092-5095.
- [11] N. Bolshakova and F. Azuaje "Cluster validation techniques" for genome expression data. *Signal Processing*, vol. 83, pp. 825-833, 2003
- [12] W. J. Conover. *Practical nonparametric statistical*. New York: John Wiley & Sons p. 428-433, 1999.
- [13] J. Murdoch, J. A. Barnes. *Statistical tables for science, engineering, business management, and finance*. 4th ed. Hampshire: Palgrave Macmillan, 1998. Table 16, pp. 28.
- [14] C. J. Mecklin and D. J. Mundfrom. "On using asymptotic critical values in testing for multivariate normality" *InterStat*, 2003. <http://interstat.stat.vt.edu/InterStat/ARTICLES/2003/articles/J03001.pdf>.
- [15] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. Canada: John Wiley & Sons, 2003, ch. 8, 11.
- [16] K. V. Mardia, A. Coombes, K. Kirkbride, A. Linney, J.L. Bowie, "On statistical problems with face identification from photographs," *J Appl Stat.*, vol. 23, no. 6, pp. 655-675, 1996.
- [17] S. R. Jammalamadaka, *Topics in circular statistics*. Singapore: World Scientific Publishing, 2001, ch. 2, 41.