Contents lists available at ScienceDirect

Forensic Science International

Forensic facial approximation assessment: can application of different average facial tissue depth data facilitate recognition and establish acceptable level of resemblance?

Lara Maria Herrera [a,c,*], Raíssa Ananda Paim Strapasson [a], Jorge Vicente Lopes da Silva [b], Rodolfo Francisco Haltenhoff Melani [a]

[a] Department of Community Dentistry, School of Dentistry, University of São Paulo, Avenida Professor Lineu Prestes, 2227, São Paulo, SP 05508-000, Brazil
[b] Three-Dimensional Technologies Division, Renato Archer Information Technology Center, Rodovia Dom Pedro I, km 143.6, Jardim Santa Mônica, Campinas, SP 13069-901, Brazil
[c] Department of Community Dentistry, Araraquara School of Dentistry, São Paulo State University, Rua Humaitá, 1680, Araraquara, SP 14801-903, Brazil

ARTICLE INFO

ABSTRACT

Facial soft tissue thicknesses (FSTT) are important guidelines for modeling faces from skull. Amid so many FSTT data, Forensic artists have to make a subjective choice of a dataset that best meets their needs. This study investigated the performance of four FSTT datasets in the recognition and resemblance of Brazilian living individuals and the performance of assessors in recognizing people, according to sex and knowledge on Human Anatomy and Forensic Dentistry. Sixteen manual facial approximations (FAs) were constructed using three-dimensional (3D) prototypes of skulls (targets). The American method was chosen for the construction of the faces. One hundred and twenty participants evaluated all FAs by means of recognition and resemblance tests. This study showed higher proportions of recognition by FAs conducted with FSTT data from cadavers compared with those conducted with medical imaging data. Targets were also considered more similar to FAs conducted with FSTT data from cadavers. Nose and face shape, respectively, were considered the most similar regions to targets. The sex of assessors (male and female) and the knowledge on Human Anatomy and Forensic Dentistry did not play a determinant role to reach greater recognition rates. It was possible to conclude that FSTT data obtained from imaging may not facilitate recognition and establish acceptable level of resemblance. Grouping FSTT data by regions of the face, as proposed in this paper, may contribute to more accurate FAs.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Facial soft tissue thicknesses (FSTT) are widely used in Forensic Sciences. Obtained in different parts of the face, they serve as guidelines for craniofacial superimposition and guide obtaining the facial contour during the process of predicting human face from skull.

Information on FSTT is often arranged in tables according to population, sex, age and body mass index [1–5]. The method of data collection is also quite varied: from needle puncture in the soft tissues of cadavers up to measurements in medical imaging [6]. The use of accurate measurements in FA techniques is very important for obtaining more real faces [7,8]. However, literature on FSTT has been mathematically and statistically analyzed, showing a lot of inconsistent data as a result of studies conducted with low scientific and methodological rigor [9–11].

Despite the trend to metrically evaluate the accuracy of FSTT mean values [6–12], it is also recommended to apply them to FAs followed by evaluation by other methods, such as recognition tests [8]. Recognition tests are methods of FA assessment that measure the ability of a target individual to be recognized from a group of people by using both familiar and unfamiliar participants [13]. Using familiar participants to assess FAs under laboratorial conditions would be ideal to simulate the performance as in real cases, but it is very difficult and unpractical [13,14]. Recognition tests are considered as good approaches of assessment of the FA accuracy [8,15].

Other possibilities of assessing FA include resemblance tests - that measures the resemblance between FA and the picture of the target during life - and methods that use a lot of images of the FA

[8,13]. According to some researchers [16,17], resemblance tests are less objective methods than recognition ones. On the other hand, a combination of all these methods could help the process of recognition, by increasing the amount of facial information details for analysis.

Aside from familiarity, students with at least a minimum of anatomical knowledge or FBI personnel have been participating as assessors of FAs in much research [13,15,16,18]. Although these studies presented different methodologies of FA sculpture and of FA assessment, recognition rates were found between 8 and 34%. Since FA is a method useful to stimulate recognition from the general public and face recognition varies considerably in the population [19], the use of different groups would be recommended for analysis, including assessors that do not have knowledge on Human Anatomy or Forensics.

Sex differences have been reported in face recognition and face perception tests, with women outperforming men [20–22]. Neurological sciences have many explanations for these findings: genetics factors [19], differences in brain physiology and also greater exposure to females during childhood [23]. However, Forensics neither has been given attention to recognition rates between male and female assessors nor to their sexual orientation, which may influence the recognition process [24].

Currently, for the Brazilian population, there are five FSTT datasets: two obtained from fresh cadavers by means of needle puncture [25,26], one obtained by MRI [27] and one obtained through computed tomography (CT) [28]. The fifth dataset refers to data obtained by means of lateral radiographs of children, so being restricted to just information of the midsagittal plane [29].

Fernandes et al. [15] have already performed computerized FAs of a Brazilian individual with FSTT standards for both Americans (cadaver data) and Brazilians (cadaver and MRI data). The authors have not obtained significant differences between recognition rates, even using 30 undergraduate students of Dentistry as examiners who were familiar with the target.

Due to the emergence of new untested FSTT data for Brazilians, this study aimed to investigate the performance of four FSTT datasets in the recognition and resemblance of individuals. It also sought to analyze the performance of assessors in recognizing people, according to sex and knowledge on Human Anatomy and Forensic Dentistry.

2. Materials and methods

2.1. Sample acquisition for FA (target individuals)

Four Brazilian volunteers (A, B, C and D) donated a CT scan and photographs of their faces, which were standardized according to authors [13,30,31] in front and profile views.

At the Renato Archer Information Technology Center (CTI), a 3D skull prototype of each target was obtained from CT images, which formed the basis for the process of modeling faces.

For making such prototypes, CT images in DICOM files were imported to the InVesalius public software (CTI, Campinas-SP, Brazil, http://www.softwarepublico.gov.br): it was set a threshold of approximately 200 HU (Hounsfield Units) in order to isolate the voxels representing bone; a mask was created and artifacts were removed; the data was rendered into a 3D model and exported as stereolitography (STL) files. In Magics software (version 1.18) from Materialise® (Belgium), a triangular mesh of each model was smoothed. For targets A and B, full-size prototypes were obtained in durable polyamide material, using a DTM Sinterstation® 2000 printer from 3D Systems® (layer thickness of 0.10 mm, laser power of 6 watts, temperature of 101 °C). For targets C and D, the full-size prototypes of the same material were produced by a Sinterstation® HiQ™ SLS® printer, from 3D Systems® (layer thickness of

0.10 mm, laser power of 18 watts, temperature of 137 °C for model C and layer thickness of 0.10 mm, laser power of 11 watts and temperature of 136 °C for model D). Then, each prototype was assessed by comparing distances between craniometrical landmarks obtained from direct measurements (using a caliper) and from measurements on CT scans (by using OsiriX software - version 5.02 35-bits, Bernex, Switzerland).

2.2. Sample of assessment tests (false targets)

Thirty-two adult volunteers (sixteen male and sixteen female) with similar characteristics (sex, age and nutritional status) to those of the targets, provided photographs of their faces for comparison with FAs, standardized according to authors [13,30,31] in front and profile views.

2.3. Sample of assessors

One hundred and twenty adult volunteers participated as FA assessors. Volunteers were divided into four groups: 30 undergraduate students who have not attended the Forensic Dentistry subject (group I); 30 undergraduate students who have attended the Forensic Dentistry subject (group II); 30 dentists who were specialists in Forensic Dentistry (group III) and 30 individuals without knowledge on Human Anatomy and Forensic Dentistry (group IV). The academic subject of Forensic Dentistry for undergraduate students included a lot of topics in Forensics, such as Forensic Anthropology and FA, not only theory contents but also demonstration and practical activities.

Some assessors casually were familiar with some targets: for example, all assessors from groups I and II were unfamiliar with targets A and B, but some of them were familiar with targets C and D; group III was unfamiliar with all individuals; some individuals from group IV were familiar with target D.

Informed consent was obtained from all participants in this study, which was approved by the Research Ethics Committee of the University of São Paulo School of Dentistry (FOUSP) under protocol of number 37709314.0.0000.0075.

2.4. Manual FA

In order to keep the blind nature of the study, the first author, who performed FAs, was not aware of the identities of the targets and false targets until completion of all FA. However, she had access to information about sex, age and nutritional status of each target.

Sculptures of the faces were conducted over the prototypes by using the American method. Four different FSTT datasets for Brazilians were also used: two datasets obtained from cadavers [25,26], one obtained from MRI [27] and one obtained from CT scans [28]. The following anatomical landmarks were excluded because they referred to points located within the soft tissue, which cannot be accurately estimated on the skull: "trichion", "supratip break", "superior labius sulcus", "stomion", "ectocanthion", "endocanthion" and "chelion". Each target had its face rebuilt four times, totaling sixteen FAs.

To compose the American method, the following parameters were used:

- Eyes: artificial eyes according to the findings of Tian et al. [cited in 32] positioned according to Krogman and Gatliff [cited in 33]: in frontal view, the top of the cornea was placed on the junction of two lines - one designed from the medial to the lateral margin of the orbit and other forming a bisector, leaving the top edge toward the bottom margin of the orbit; in lateral view, the tip of the cornea was at the center of a line drawn between the top and lower margins, so that to form a tangent projected 2 mm [33].

- Filling: clay strips were used to filling the gaps between stops of FSTT over the skull. The procedure began across the frontal bones, followed to the masseter muscle area and ended on the chin [33].
- Mouth: mouth was built using 3 dimensions: depth, height and width. The depth of lips was determined by the "infra-dentale" tissue depth marker; the height was obtained by measuring the distance between the lines of the cementum-enamel junction of the upper and lower central incisors; mouth width was determined by measuring the distance between the distal surfaces of canine teeth (measured in arc) according to Krogman [cited by 33,34]. Finally, the parting line was incised at the halfway point of the vertical thickness of lips [33].
- Nose: In a frontal view, the nasal aperture was measured at its widest point and 10 mm was added for Caucasoids (5 mm to each side) to get the total width. In a lateral view, to estimate the nasal projection, the length of the anterior nasal spine was obtained and multiplied by 3. This value was added to the depth of tissue depth marker "P5 – MidPhiltrum". The direction and morphology of the anterior nasal spine were also used to sculp the shape of the nose [33].
- Ears: ear length was 50 mm and ears were positioned behind the mandibular ramus [Krogman cited by 35].
- Neck: it was modeled with soft clay on the support rod.
- Texturing and characterization: a thin layer of clay was applied on the face for covering the stops, so that they would not interfere on face shape and tissue thickness. A sponge was applied to create the skin pores and facial features were built, based on the age and sex of each target. Eyebrows were also sculpted according to Stephan [36] and a standard hairstyle for females and another for males were adopted.

When the sculpture was finished, photographs in front and profile standards were taken.

FAs received numbers: FA1, FA2, FA3 and FA4 represented FAs of target A using the four FSTT datasets: cadaver [25] cadaver [26], MRI [27] and CT [28], respectively; FA5 to FA8 represented FAs of target B; FA9 to FA12 represented FAs of target C; and FA13 to FA16 represented FAs of target D.

FAs were constructed following a randomized sequence for both models and FSTT data. The sequence was obtained by means of the website "http://www.random.org".

2.5. FA assessment

FA assessment was conducted by means of recognition and resemblance tests.

For recognition test, each FA was presented among a face pool of nine faces (target and 8 false targets) simultaneously and assessors were asked to recognize the target. They could also choose "none" when they believed that the target was not in the face pool.

For the resemblance test, each FA was shown close to the target (side by side, to verify the level of resemblance between them), as well as to other two false targets, sequentially, to compose a fake test. It was used 3 arrays instead of 9 to not become the assessment troublesome and time-intensive, as already stated for some authors [16]. Assessors had to indicate the degree of resemblance between FA and faces. Resemblance levels were defined according to Stephan and Cicolini [16] and arranged on a scale from 1 to 5, as follows: 1 = no resemblance; 2 = slight resemblance; 3 = approximate; 4 = close; 5 = strong resemblance. When minimal resemblance was chosen (score 2), the volunteers pointed out the face regions considered of most resemblance (face shape, forehead, eyes, nose, mouth, chin, cheeks, and/or ears).

Assessments were conducted with groups I, II and IV on-site, separately. For group III, evaluation was conducted remotely, via online forms.

2.6. Statistical analysis

2.6.1. Recognition test

Relative frequencies (%) of correct faces (targets) were obtained per FSTT dataset. The frequencies of success were also observed in relation to sex and group of assessors (chi-squared tests or Fisher's exact test [37]).

The binomial Generalized Estimating Equations (GEE) model [38] was also used, with the main purpose of considering the effect of interaction between FSTT dataset and group of assessors.

2.6.2. Resemblance test

The absolute (n) and relative (%) frequencies of some resemblance (at least score 2) were observed and the comparison among the resemblance rates by the four FAs was conducted by paired comparison test for proportions.

The frequencies of the face regions indicated as of higher resemblance were also obtained (score > 1).

To consider the dependence among observations of the same assessor and to check if there is effect of interaction between FSTT data and group of assessors, i.e., if the difference of resemblance scores between FAs (FSTT data) vary (or differ) in these groups of assessors, a non-parametric analysis of ordinal data for repeated measures was performed [39].

The level of significance was 5%. p-Values greater than 0.05 and less than 0.10 were also considered, suggesting difference with weaker evidence. The software used were: SPSS (Statistical Package for Social Sciences) version 19; R version 3.1.1; macro excel in: http://www.ime.usp.br/~jmsinger/; Microsoft Excel® and Word®.

3. Results

Fig. 1 shows all FAs of the four targets (A, B, C and D) constructed with four different FSTT datasets for Brazilians.

All FAs were evaluated by a hundred and twenty volunteers aged 18–73 years, mean age of 31 years (SD = 6.12 years), 83 (69.2%) were female and 37 (30.8%) were male.

3.1. Recognition test:

Fig. 2 shows the success (correct recognition) per FSTT dataset. Table 1 shows the frequencies of correct recognition in each category and the results of the tests of association between sex of the assessors and success recognition rates.

The results of the tests of association (Table 1) suggest that only FA7 (FA of target B, male, by MRI data) shows significant evidence of a greater proportion of success by men (32%) than by women (15%).

To find out the behavior of the differences between the proportions of success recognition in relation to groups, considering the dependence of observations, the effect of the interaction between FAs (FSTT datasets) and group was evaluated in the logistics GEE model. The results are shown in Table 2.

The results from the GEE models (Table 2), in the presence of effects of interaction, suggest significant effects of FSTT dataset for all targets. It was also verified significant effects of group for targets A, B and C and significant effect of interaction between group and FSTT dataset for target C.

Although not shown in this paper, after multiple comparisons tests, it was possible to locate where these effects happened and also to know how much the difference of proportions of success was. Regarding to effects of FSTT data, for target A (female), FA conducted with tables of cadavers presented higher proportions of correct responses than those conducted with MRI and CT data. In
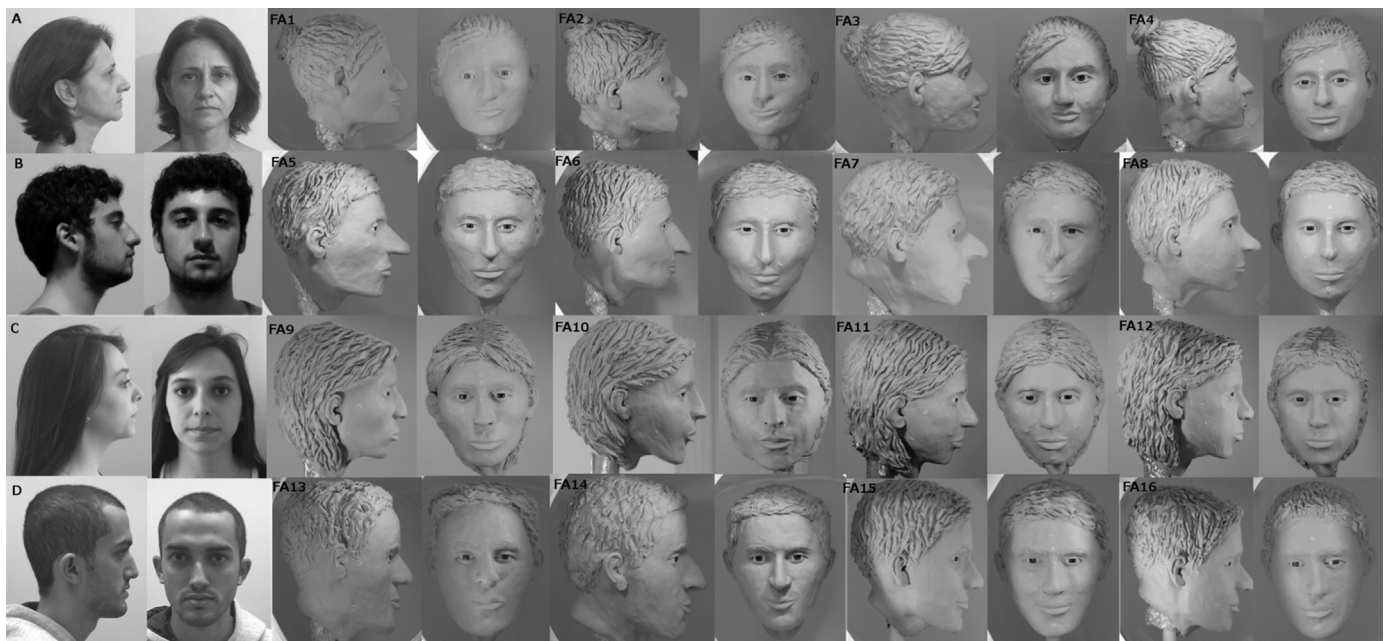
Fig. 1. Facial approximations of the targets (A, B, C and D). FA1, FA5, FA9 and FA13 were conducted with FSTT data from cadaver proposed by Tedeschi-Oliveira [25]; FA2, FA6, FA10 and FA14 were conducted with FSTT data from cadaver proposed by Almeida et al. [26]; FA3, FA7, FA11 and FA15 were conducted with FSTT data from MRI proposed by Santos [27]; FA4, FA8, FA12 and FA16 were conducted with FSTT data from CT proposed by Beaini [28].
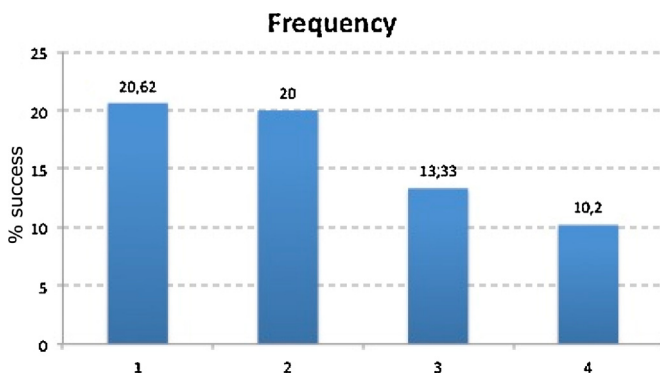
Fig. 2. Relative frequencies of success per FSTT dataset. 1: cadaver proposed by Tedeschi-Oliveira et al. [25]; 2: cadaver proposed by Almeida et al. [26]; 3: MRI proposed by Santos [27]; 4: CT proposed by Beaini [28].

this case, FA with MRI data showed the lowest proportion of success. For target B (male), FA with cadaver data of Almeida et al. [26], MRI and CT showed, on average, higher proportions of success than FA with cadaver data of Tedeschi-Oliveira et al. [25]. For target C (female), FA conducted with cadaver data from Tedeschi-Oliveira et al. [25] presented the lowest proportions of success compared with those conducted with MRI and CT data. For target D (male), FA conducted with CT data showed lower proportion of success than FA with cadavers' data.

Regarding to effects of groups, behaviors were found to be varied, in general. For target A, group I presented smaller proportions of correct responses (correct recognition) than the other groups. For target B, group II showed lower proportions of success compared to others. For target C, only group II showed differences in the proportions of success among FA. In FA with CT data, group II showed higher success rates than groups I and III. For target D, it was not observed effects of group.

3.2. Resemblance test

The resemblance assessment was performed for the target and for two false targets (a, b, c). Table 3 shows the frequency distribution of at least slight resemblance (score > 1) and the significant differences in the proportions of slight resemblance among FAs.

Fig. 3 shows the frequency of some resemblance (at least slight resemblance) between FAs and targets per FSTT dataset.

Table 3 shows that the proportion of some resemblance (score > 1) did not differ significantly among FAs with different FSTT datasets for faces of targets B and D (male) and target C (female). For the face of target A (female), FA with cadaver data of Tedeschi-Oliveira et al. [25] presented higher proportion (88%) of some resemblance in relation to FAs with CT and MRI data (less than 70%).

The frequencies of regions considered of greater resemblance were obtained only for cases where there was at least some resemblance, and more than one region, among eight, could be singled out (Table 4).

Table 4 shows that nose and face shape, in that order, were regions considered of greatest resemblance for FSTT datasets of Tedeschi-Oliveira et al. [25], Santos [27] and Beaini [28]. For FAs conducted with the dataset of Almeida et al. [26], the second most frequent region was the chin.

To evaluate the behavior of the differences among resemblance scores among FAs, both for female and male targets, in relation to groups, considering the dependence of observations among methods (FSTT datasets), a nonparametric model of ordinal data for repeated measures was adjusted and the results are shown in Table 5.

The analysis of ordinal data shows the effect of interaction between group and FA for all targets, although with weaker evidence for target A (female) ($0.05 < p < 0.10$). These results emphasize the different behavior of resemblance scores with respect to FA for each group, and each FA may also differ in the resemblance scores in relation to different groups.

Table 1
Associations between sex and success recognition rates.

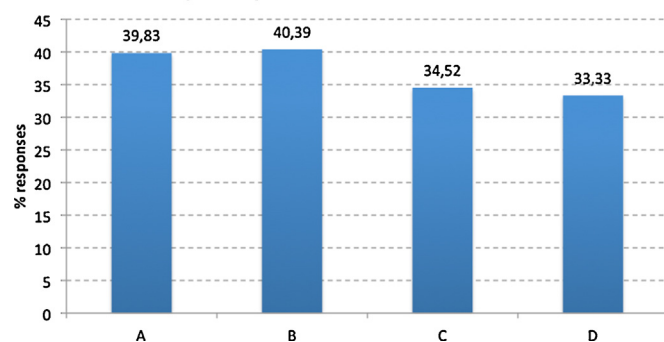|		Sex		p-Value			Sex		p-Value
		F	M				F	M	
FA1	n	41	12		FA9	n	2	3	
	%	49.4%	32.4%	**0.084[P]**		%	2.4%	8.1%	0.170[F]
	SR	1.7	−1.7			SR	−1.4	1.4	
FA2	n	35	11		FA10	n	7	4	
	%	42.2%	29.7%	0.196[P]		%	8.4%	10.8%	0.677[P]
	SR	1.3	−1.3			SR	−0.4	0.4	
FA3	n	5	5		FA11	n	3	1	
	%	6.0%	13.5%	0.170[P]		%	3.6%	2.7%	1.000[F]
	SR	−1.4	1.4			SR	0.3	−0.3	
FA4	n	12	5		FA12	n	5	1	
	%	14.5%	13.5%	0.891[P]		%	6.0%	2.7%	0.665[F]
	SR	0.1	−0.1			SR	0.8	−0.8	
FA5	n	8	4		FA13	n	19	10	
	%	9.6%	10.8%	0.843[P]		%	22.9%	27.0%	0.625[P]
	SR	−0.2	0.2			SR	−0.5	0.5	
FA6	n	16	3		FA14	n	17	3	
	%	19.3%	8.1%	0.122[P]		%	20.5%	8.1%	**0.093[P]**
	SR	1.5	−1.5			SR	1.7	−1.7	
FA7	n	12	12		FA15	n	19	7	
	%	14.5%	32.4%	**0.023[P]**		%	22.9%	18.9%	0.626[P]
	SR	−2.3	2.3			SR	0.5	−0.5	
FA8	n	12	4		FA16	n	6	4	
	%	14.5%	10.8%	0.587[P]		%	7.2%	10.8%	0.512[P]
	SR	0.5	−0.5			SR	−0.7	0.7	

P: Pearson's chi-square test, F: Fisher's exact test. SR: standardized residual (excess and lack in the category can be assigned to SR values greater than 1.96 and less than −1.96, respectively).

Table 2
Tests of effects in the GEE model.

	Wald statistics (Chi-square)	Degrees of freedom	p-Value
Model for Target A (female)			
Intercept	508.8	1	**<0.001**
Group	290.4	3	**<0.001**
FSTT dataset	320.5	3	**<0.001**
Group*FSTT dataset	9.4	8	0.307
Model for Target B (male)			
Intercept	559.8	1	**<0.001**
Group	351.3	3	**<0.001**
FSTT dataset	305.8	3	**<0.001**
Group*FSTT dataset	8.8	8	0.357
Model for Target C (female)			
Intercept	715.0	1	**<0.001**
Group	508.5	3	**<0.001**
FSTT dataset	339.9	3	**<0.001**
Group*FSTT dataset	483.2	4	**<0.001**
Model for Target D (male)			
Intercept	161.0	1	**<0.001**
Group	1.4	3	0.700
FSTT dataset	10.5	3	**0.015**
Group*FSTT dataset	15.0	9	**0.091**

Fig. 3. Frequency of at least some resemblance (score > 1) between FAs and targets. A: cadaver FSTT dataset proposed by Tedeschi-Oliveira et al. [25]; B: cadaver FSTT proposed by Almeida et al. [26]; C: MRI FSTT dataset proposed by Santos [27]; D: CT FSTT dataset proposed by Beaini [28].

4. Discussion

This study showed higher proportions of recognition for targets by FAs constructed with cadaver FSTT datasets in relation to medical imaging datasets (Fig. 2). By analyzing only the frequencies of some resemblance (score > 1), it was observed that targets were considered more similar to FAs conducted with cadaver FSTT data too (Fig. 3). Only target A showed significant differences of frequencies of slight resemblance among FAs and FA with cadaver data of Tedeschi-Oliveira et al. [25] presented the highest proportion of some resemblance in relation to FAs with MRI and CT data (Table 3).

Fernandes et al. [15] did not conclude whether there is a FSTT dataset that can be consider the best one to use, since at least one correct recognition was obtained for all FAs with different FSTT data used in their study (including an "international pattern"). The authors, however, used only thirty examiners and suggested repeating the study with a larger sample. Stephan and Cicolini [14] recommend teams above hundred and fifteen individuals, otherwise, the results should be considered approximate. This assessment used one hundred and twenty examiners and also obtained at least one correct recognition for each case. We did not found other studies performing FAs with different FSTT datasets of the same population for comparison.

It could be concluded, therefore, that considering FA with FSTT data obtained by imaging, such as CT and MRI, as the most accurate is a mistake. Although some studies [6,40] claim that medical imaging is the best method for obtaining FSTT values due to the possibility of locating anatomical landmarks of interest for

Table 3
Frequency distribution of at least slight resemblance (score >1) and identification of significant differences among FAs.

	a		b		c		Total	
	n	%	n	%	n	%	n	%
FA1	67	55.8**	106	88.3**	64	53.3*	237	65.8**
FA2	34	28.3*	94	79.0	47	39.17*	175	48.7*
FA3	55	45.8**	80	67.2*	93	77.5**	228	63.5**
FA4	76	63.3**	82	68.3*	49	40.8*	207	57.5
FA5	56	46.7	25	20.8*	34	28.3*.	115	31.9*
FA6	73	60.8	39	32.5	48	40.0.	160	44.4**..
FA7	72	60.0	55	45.8**	78	65.0**.**	205	56.9**.**
FA8	67	55.8	52	43.3**	65	54.2**.	184	51.1**
FA9	37	30.8*	29	24.2*	52	43.3	118	32.8*
FA10	85	70.8***	77	64.2**	46	38.3	208	57.8**
FA11	58	48.3**	74	62.2**	53	44.2	185	51.5**
FA12	94	79.0**	85	71.4**	59	49.2	238	66.5**
FA13	80	66.7***	74	61.7	99	82.5**	253	70.3***
FA14	21	17.5*	73	60.8	71	59.2*	165	45.8*
FA15	49	40.8**	76	63.3	71	59.2*	196	54.4**
FA16	48	40.0**	57	47.5	57	47.5*	162	45.0*

p-Value among approximations: paired comparisons among proportions.
Significant differences between pairs of approximations: ***> **> * (*** showed values significantly higher than ** and ** values significantly higher than*)

Table 4
Frequencies of the regions of the face of greatest resemblance in cases where there was at least some resemblance (score > 1).

Regions of the face	Cadaver (Tedeschi-Oliveira et al. [25])		Cadaver (Almeida et al. [26])		MRI (Santos [27])		CT (Beaini [28])	
	N	%	N	%	N	%	N	%
Nose	160	55.56	141	49.30	125	44.48	113	42.64
Face shape	97	33.68	90	31.47	117	41.64	125	47.17
Forehead	94	32.64	57	19.93	67	23.84	79	29.81
Eyes	88	30.56	84	29.37	101	35.94	85	32.08
Chin	84	29.17	112	39.16	73	25.98	91	34.34
Mouth	81	28.13	106	37.06	93	33.10	78	29.43
Cheeks	49	17.01	42	14.69	51	18.15	55	20.75
Ears	25	8.68	32	11.19	40	14.23	30	11.32

Table 5
Tests of the effects on the non-parametric model of ordinal data.

	Wald Statistics	D.f.	*p*-Value
Target A (female)			
Group	2.32	3	0.509
FSTT dataset	31.82	3	**<0.001**
Interaction (Group*FSTT dataset)	16.60	9	**0.055**
Target B (male)			
Group	16.31	3	**0.001**
FSTT dataset	8.24	3	**0.041**
Interaction (Group*FSTT dataset)	23.26	9	**0.006**
Target C (female)			
Group	4.67	3	0.198
FSTT dataset	5.69	3	0.128
Interaction (Group*FSTT dataset)	23.69	9	**0.005**
Target D (male)			
Group	6.28	3	0.099
FSTT dataset	15.95	3	**0.001**
Interaction (Group*FSTT dataset)	19.74	9	**0.020**

measurement, there are many limitations attributed to the technique. As pointed out by other authors [41,42], the measurement direction, the head positioning, the resolution of images, the segmentation algorithms of hard tissues and positioning of markers are factors that can interfere in data collection. Thus either using cadaver or medical imaging examination data, each method has advantages and limitations.

On the other hand, some studies [9–11] have recently pointed out that the division of FSTT data by methods of obtainment is of little practical significance, as well as the division of data by sex,

age, population and body mass index. The main reason for such observations is the low methodological rigor in the collection of these thicknesses, such as sample size and measurement errors.

Stephan and Simpson [6] compared FSTT data obtained by different methodologies and found that the averages obtained by needle puncture in the soft tissues of cadavers and ultrasound were quite similar and the averages obtained by CT were the lowest ones compared to all other methods; however, the authors used a different sample of the CT data for analysis. The authors also observed similar patterns of average values for men and women and conclude that the protocols for obtaining FSTT are quite varied, even for the same method of collection. They suggested the use of a single FSTT dataset without divisions into sub-categories (also known as Taillied-Facial-Soft-Tissue-Depth-Data or T-Table) created by them in 2008 and updated in 2013 due to the increased number of data in literature [11]. Guyomarc'h et al. [4] found no differences between FSTT data of individuals of different sexes, population and ages and suggest the application of textures (hair, wrinkles and makeup) to improve the visual impact of FA than solely based on FSTT to obtain a more accurate FA. The FSTT dataset of cadavers obtained by Almeida et al. [26] did not divide the average thicknesses by sex and its performance in this study can be considered satisfactory: FA constructed with this dataset allowed more correct recognition in relation to FA constructed with CT dataset for all targets, although significant only for target A. According to the authors [26], the minimum differences of FSTT between men and women may be associated to face shape instead of sex. Further research could be conducted to verify, for example, if FAs using datasets of unified data, such as the T-table, enables better results. Lee et al. [8], however, claim that such assessments should be conduct by recognition tests.

Based on data obtained by association tests, the sex of assessors (male and female) did not play a determinant role to reach greater recognition rates, since only FA7 showed significant evidence of greater proportion of success by men compared to women (Table 1). It was expected to find targets being statistically more recognized by female assessors, because according to some studies [20–22,43], women are more skilled at recognizing faces than the opposite sex.

In addition, women tend to better recognize faces of women and men, faces of men [20–22]. This fact is explained because women are possibly more familiar with or have more knowledge of female faces than male faces as a result of greater exposure to beauty magazines or due to the different interests between sexes [20,21]. It is interesting that the frequency of correct recognition related to female targets were higher compared to male targets, which reinforces the idea presented above, since most assessors of this study were female (69%). However, others experimental neuropsychological studies have been appointing opposite sex bias for men in face recognition [23] or inconclusive results [24]. A next approach could be taking into account the sexual orientation of the assessors, which might have influence in face recognition.

Among groups of assessors, a quite varied behavior of recognition and resemblance was observed in relation to the different FSTT datasets, and it was not possible to conclude that individuals with greater knowledge on Human Anatomy or Forensic Dentistry have higher performance in recognizing people. This finding strengthens the aim of FA, which is to benefit the whole society, that is, any individual who has a missing relative or friend in need of identification may be able to recognize him or her. It was not possible to compare the results of this research with others in literature, which usually uses assessors with some anatomical knowledge.

Unlike what is pointed out in literature, this study found that being a familiar assessor with the target not necessarily provided higher frequencies of correct recognitions and greater resemblance scores. All groups was unfamiliar with targets A and B and differences in the proportions of correct recognition were observed: for target A, only group I showed lower success rate compared to the other groups and for target B, group II had the lowest proportions. Some assessors of groups I, II and IV were familiar with target D and no effect of group on the differences of correct proportions was observed. There was an effect of group on the resemblance scores only for target B. One possible explanation for the results is the fact that some assessors were *visually* not *personally* familiar with some targets. Personal familiarity is related to close people like parents, children and friends, so the amount of information involved is different (not only about facial appearance but also about personal traits, emotions and so on) [44].

Some authors [13,14,45] argue that higher recognition rates in research are obtained from assessors familiar with targets but very few studies have managed to use them. Stephan et al. [46] performed 2D FAs that were assessed by unfamiliar and familiar groups of assessors. The results with familiar assessors were considered more robust; however, the sample of familiar assessors was very small and one of the targets had very specific facial features compared to the others, which may have increased recognition frequencies. Fernandes et al. [15] used only familiar assessors (thirty individuals) when assessing 3D FAs and obtained frequencies of success that did not exceed 24%. In the present study, although the distribution of assessors was not homogeneous regarding familiarity, the recognition frequencies for female targets, together, did not exceed 27%, being higher than the frequencies obtained by Fernandes et al. [15]. For male targets, the recognition frequencies, together, did not exceed 23%.

Although not shown in the results, it was verified for all FAs that the greater the number of face regions identified as of major resemblance, the greater the resemblance scores (Pearson correlation test, $p < 0.001$). From the results obtained in Table 4 and because studies follow a trend towards unification of FSTT

Table 6
Table of facial soft tissue thicknesses for Brazilian subjects.

Facial regions	Anatomical landmarks	FSTT data	Mean (mm)		
			Male	Female	Both sexes
Nose	End of nasals	Cadaver (Tedeschi-Oliveira et al., 2009)	x	x	4.7
	Midphiltrum	Cadaver (Tedeschi-Oliveira et al. 2009)	10.6	7.7	x
	Nasion	Cadaver (Tedeschi-Oliveira et al. 2009)	x	x	5.4
Face	Zigomatic arch	CT (Beaini, 2013)	x	x	7.7
	Gonion (R;L)	CT (Beaini, 2013)	17.3	13.2	x
Eyes	Supraorbital	MRI (Santos, 2008)	x	x	8.7
	Suborbital	MRI (Santos, 2008)	x	x	6.3
	Lateral orbit	MRI (Santos, 2008)	x	x	9.4
Forehead	Supraglabella	Cadaver (Tedeschi-Oliveira et al., 2009)	x	x	4.6
	Glabella	Cadaver (Tedeschi-Oliveira et al., 2009)	5,5	4,6	x
	Frontal eminence	Cadaver (Tedeschi-Oliveira et al., 2009)	4.9	3.9	x
Chin	Chin-lipfold	Cadaver (Almeida et al., 2013)	x	x	9.92
	Mental eminence	Cadaver (Almeida et al., 2013)	x	x	11.04
	Beneathchin	Cadaver (Almeida et al., 2013)	x	x	9.78
Mouth	Supradentale	Cadaver (Almeida et al., 2013)	x	x	8.18
	Infradentale	Cadaver (Almeida et al., 2013)	x	x	7.58
	Supra-canine	Cadaver (Almeida et al., 2013)	x	x	13.62
	Sub-canine	Cadaver (Almeida et al., 2013)	x	x	12.58
Cheeks	Supra-M2	CT (Beaini, 2013)	28.7	26.1	x
	Infra-M2 (R;L)	CT (Beaini, 2013)	25.12; 25.53	23.52; 23.99	x
	Occlusal line	CT (Beaini, 2013)	22.8	20.2	x
	Inferior malar	CT (Beaini, 2013)	x	x	19,7
Ears	Supraglenoid	MRI (Santos, 2008)	14.26	13.23	x

R: Right; L: Left.

data, this study proposed to group the four FSTT datasets, by region, as seen in Table 6.

It was observed that the mouth region of the FAs was more selected as similar to the mouth of the targets when using the cadaver dataset of Almeida et al. [26]. This might have occurred because the authors proposed the inclusion of two additional craniometric landmarks (canines) for mouth, which supposedly have contributed to its better configuration in the sculpture.

Another notable observation can be made for the regions of the cheeks. In this region, there are many fatty structures such as the corpus adiposum buccae, also known as "buccal fat pad of Bichat" [47]. The size of such tissue varies among subjects due to body weight difference and therefore it influences the FSTT in the region of cheeks [5]. The impact on FAs conducted with data from cadavers, for these regions, is a leaner face with deepened aspects of the tissues, which was not observed for FA conducted with CT data, enabling a more natural contour of the cheeks. This was, probably, due to the higher thicknesses values obtained in these regions with medical imaging, especially those from CT, compared to measurements obtained from cadavers.

It is important to consider the limitations of this study. FA is an additional Forensic resource that requires not only subjective interpretation of the skull but also artistical abilities, regardless of the method chosen (manual or computerized). Although the researcher that conducted all FA have attended a specific workshop on Forensic sculpture, the subjectivity and her experience may have had impact on the repeated approximations of the same skull. It is very difficult for Forensic Artists to reproduce exactly the same facial structures. It was evident, for example, that eyes, nose, mouth and ears presented different shapes among FAs of the same target, even using the same technique to build these structures. The varied stops of FSTT data, on the other hand, were glued over the same exactly anatomical landmarks for the same target (we located and marked the points over each skull with a pen just once). However, the variations observed cannot only be assigned to a change of FSTT data in those regions but to variations that are unavoidable in manual FA, which may add some bias in the recognition process. Another limitation is that the forms given to the assessors did not include a specific field related to the familiarity, so the sample was not representative regarding this aspect.

We believe the findings of this research, especially those related to the development of the guideline of FSTT (Table 6), help to elucidate the uncertainty that the rebuilder might have facing different FSTT data for Brazilians. We also believe these results support the need for further exploration of the impact of different FSTT data versus pooled FSTT means (such as the T-Table and the presented guideline for Brazilians) on the recognition process.

5. Conclusions

Given the overall results from this study, it is possible to conclude that FSTT data obtained from imaging may not facilitate recognition and establish acceptable level of resemblance in FFA. From the FA assessments, it was possible to create a combined data of the mainly FSTT datasets for Brazilians and the greatest gain to the scientific community is the possibility of using the best benefits that each table might offer to reach a more satisfactory FA. Furthermore, this research emphasizes the need for investigation of other factors, different from sex and expertise of the assessors that can impact on face recognition.

Acknowledgements

References

[1] P. Panenkova, R. Benus, S. Masnicova, Z. Obertova, J. Grunt, Facial soft tissue thicknesses of the mid-face for Slovak population, Forensic Sci. Int. 220 (1–3) (2012), 293.e1-e6.

[2] S. De Greef, P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, G. Willems, Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction, Forensic Sci. Int. 159 (Suppl. 1) (2006) S126–S146.

[3] W. Jeelani, M. Fida, A. Shaikh, Facial soft tissue thickness among three skeletal classes in adult Pakistani subjects, J. Forensic Sci. (2015), http://dx.doi.org/10.1111/1556-4029.12851.

[4] P. Guyomarc'h, F. Santos, B. Dutailly, H. Coqueugniot, Facial soft tissue depths in French adults: variability, specificity and estimation, Forensic Sci. Int. 231 (1–3) (2013), 411.e1-e10.

[5] S. De Greef, D. Vandermeulen, P. Claes, P. Suetens, G. Willems, The influence of sex, age and body mass index on facial soft tissue depths, Forensic Sci. Med. Pathol. 5 (2) (2009) 60–65.

[6] C.N. Stephan, E.K. Simpson, Facial soft tissue depths in craniofacial identification (part I): an analytical review of the published adult data, J. Forensic Sci. 53 (6) (2008) 1257–1272.

[7] W.J. Lee, C.M. Wilkinson, H.S. Hwang, An accuracy assessment of forensic computerized facial reconstruction employing cone-beam computed tomography from live subjects, J. Forensic Sci. 57 (2) (2012) 318–327.

[8] W.J. Lee, C.M. Wilkinson, H.S. Hwang, S.M. Lee, Correlation between average tissue depth data and quantitative accuracy of forensic craniofacial reconstructions measured by geometric surface comparison method, J. Forensic Sci. 60 (3) (2015) 572–580.

[9] C.N. Stephan, Accuracies of facial soft tissue depth means for estimating ground truth skin surfaces in forensic craniofacial identification, Int. J. Leg. Med. 129 (4) (2015) 877–888.

[10] C.N. Stephan, L. Munn, J. Caple, Facial soft tissue thicknesses: noise, signal, and P, Forensic Sci. Int. 257 (2015) 114–122.

[11] C.N. Stephan, The application of the central limit theorem and the law of large numbers to facial soft tissue depths: T-Table robustness and trends since 2008, J. Forensic Sci. 59 (2) (2014) 454–462.

[12] C.N. Stephan, E.K. Simpson, Facial soft tissue depths in craniofacial identification (part II): an analytical review of the published sub-adult data, J. Forensic Sci. 53 (6) (2008) 1273–1279.

[13] A.H. Richard, C.L. Parks, K.L. Monson, Assessment of presentation methods for ReFace computerized facial approximations, Forensic Sci. Int. 242 (2014) 283–292.

[14] C.N. Stephan, J. Cicolini, The reproducibility of facial approximation accuracy results generated from photo-spread tests, Forensic Sci. Int. 201 (1–3) (2010) 133–137.

[15] C.M. Fernandes, C. Serra Mda, J.V. da Silva, P.Y. Noritomi, F.D. Pereira, R.F. Melani, Tests of one Brazilian facial reconstruction method using three soft tissue depth sets and familiar assessors, Forensic Sci. Int. 214 (1–3) (2012), 211.e1-e7.

[16] C.N. Stephan, J. Cicolini, Measuring the accuracy of facial approximations: a comparative study of resemblance rating and face array methods, J. Forensic Sci. 53 (1) (2008) 58–64.

[17] C.N. Stephan, R.S. Arthur, Assessing facial approximation accuracy: how do resemblance ratings of disparate faces compare to recognition tests? Forensic Sci. Int. 159S (2006) S159–S163.

[18] C.N. Stephan, M. Henneberg, Building faces from dry skulls: are they recognized above chance rates? J. Forensic Sci. 46 (3) (2001) 432–440.

[19] S. Karlsson, S. Henningsson, D. Hovey, A. Zettergren, L. Jonsson, D.S. Cortes, J. Melke, P. Laukka, H. Fischer, L. Westberg, Social memory associated with estrogen receptor polymorphisms in women, Soc. Cogn. Afffect. Neurosci. (March) (2016), http://dx.doi.org/10.1093/scan/nsw010.

[20] C. Lewin, A. Herlitz, Sex differences in face recognition – women's faces make the difference, Brain Cogn. 50 (1) (2002) 121–128.

[21] J. Rehnman, A. Herlitz, Women remember more faces than men do, Acta Psychol. (Amst.) 124 (3) (2007) 344–355.

[22] D.B. Wright, B. Sladden, An own gender bias and the importance of hair in face recognition, Acta Psychol. (Amst.) 114 (1) (2003) 101–114.

[23] M. Verdichevski, J.K.E. Steeves, Own-age and own-sex biases in recognition of aged faecs, Acta Psychol. 144 (2) (2013) 418–423.

[24] M.C. Steffens, S. Landmann, S. Mecklenbräuker, Participant sexual orientation matters: new evidence on the gender bias in face recognition, Exp. Psychol. 60 (5) (2013) 362–367.

[25] S.V. Tedeschi-Oliveira, R.F. Melani, N.H. de Almeida, L.A. de Paiva, Facial soft tissue thickness of Brazilian adults, Forensic Sci. Int. 193 (1–3) (2009), 127.e1-e7.

[26] N.H. de Almeida, E. Michel-Crosato, L.A. de Paiva, M.G. Biazevic, Facial soft tissue thickness in the Brazilian population: new reference data and anatomical landmarks, Forensic Sci. Int. 231 (1–3) (2013), 404.e1-e7.

[27] W.D.F. Santos, Mensuração de tecidos moles da face de brasileiros vivos em imagens multiplicares de Ressonância Magnética Nuclear (RMN) para fins médico-legais, Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, 2008.

[28] T.L. Beaini, Espessura de tecidos moles nos diferentes tipos faciais: estudo em tomografias computadorizadas cone-beam, Universidade de São Paulo, Faculdade de Odontologia, São Paulo, 2013.

[29] M.M. Pithon, D.L. Rodrigues Ribeiro, R. Lacerda dos Santos, C. Leite de Santana, J.P. Pedrosa Cruz, Soft tissue thickness in young north eastern Brazilian individuals with different skeletal classes, J. Forensic Leg. Med. 22 (2014) 115–120.

[30] S. Hayes, R. Taylor, A. Paterson, Forensic facial approximation: an overview of current methods used at the Victorian Institute of Forensic Medicine/Victoria Police Criminal Identification Squad, J. Forensic Odontostomatol. 23 (2) (2005) 45–50.

[31] G. Quatrehomme, T. Balaguer, P. Staccini, V. Alunni-Perret, Assessment of the accuracy of three-dimensional manual craniofacial reconstruction: a series of 25 controlled cases, Int. J. Leg. Med. 121 (6) (2007) 469–475.

[32] C.M. Wilkinson, S.A. Mautner, Measurement of eyeball protrusion and its application in facial reconstruction, J. Forensic Sci. 48 (1) (2003) 12–16.

[33] K.T. Taylor, Forensic Art and Illustration, CRC Press, New York, 2001.

[34] C. Wilkinson, Facial approximation: comments on Stephan (2003), Am. J. Phys. Anthropol. 125 (4) (2004) 329–330.

[35] P. Guyomarc'h, C.N. Stephan, The validity of ear prediction guidelines used in facial approximation, J. Forensic Sci. 57 (6) (2012) 1427–1441.

[36] C.N. Stephan, Position of superciliare in relation to the lateral iris: testing a suggested facial approximation guideline, Forensic Sci. Int. 130 (2002) 29–33.

[37] P. Armitage, G. Berry, Statistical Methods in Medical Research, Blackwell Science, Oxford, 1994.

[38] K.Y. Liang, S.L. Zeger, Longitudinal analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[39] E. Brunner, F. Langer, Nonparametric analysis of ordered categorical data in designs with longitudinal observations and small sample sizes, Biometr. J. 42 (2000) 663–675.

[40] N.A. Perlaza Ruiz, Facial soft tissue thickness of Colombian adults, Forensic Sci. Int. 229 (1–3) (2013), 160. e1-e9.

[41] H.S. Hwang, S.Y. Choe, J.S. Hwang, D.N. Moon, Y. Hou, W.J. Lee, et al., Reproducibility of facial soft tissue thickness measurements using cone-beam CT images according to the measurement methods, J. Forensic Sci. 60 (4) (2015) 957–965.

[42] J.M. Caple, C.N. Stephan, L.S. Gregory, D.M. MacGregor, Effect of head position on facial soft tissue depth measurements obtained using computed tomography, J. Forensic Sci. (2015) http://dx.doi.org/10.1111/1556-4029.12896.

[43] J. Rehnman, A. Herlitz, Higher face recognition ability in girls: magnified by own-sex and own-ethnicity bias, Memory 14 (3) (2006) 289–296.

[44] V. Natu, A.J. O'Toole, The neural processing of familiar and unfamiliar faces: a review and synopsis, Br. J. Psychol. 102 (2011) 726–747.

[45] S. Davy-Jow, The devil is in the details: a synthesis of psychology of facial perception and its applications in forensic facial reconstruction, Sci. Justice 53 (2) (2013) 230–235.

[46] C.N. Stephan, I.S. Penton-Voak, J.G. Clement, M. Henneberg, Ceiling recognition limits of two-dimensional facial approximations constructed using averages, in: J.G. Clemente (Ed.), Computer Graphic Facial Reconstruction, Academic Press, Boston, 2005, pp. 199–219.

[47] P.M. Tostevin, H. Ellis, The buccal pad of fat: a review, Clin. Anat. 8 (6) (1995) 403–406.