The use of three-dimensional surface scans for facial feminization surgery

Quantification of the indications and results

H.H.G. Dankers
Technical Medicine - Medical Imaging & Interventions - University of Twente

3D Lab, Radboudumc Nijmegen
Department of Oral and Maxillofacial Surgery, Amsterdam UMC - VUmc

April 4th, 2023

Supervisors

Prof. dr. I. Sechopoulos	Chairman & Technical supervisor, University of Twente
Drs. B.A. Meijer	Medical supervisor, Amsterdam UMC - VUmc
Prof. dr. E.A.J.M. Schulten	Medical supervisor, Amsterdam UMC - VUmc
Dr. G.A. de Jong	Technical supervisor, Radboudumc Nijmegen
Prof. dr. T.J.J. Maal	Technical supervisor, Radboudumc Nijmegen
Drs. P.A. van Katwijk	Process supervisor, University of Twente
Dr. ir. L. Alic	External member, University of Twente

Abstract

Facial feminization surgery (FFS) comprises a broad range of craniomaxillofacial surgical procedures with the objective to change masculine facial features into feminine features. FFS can play an important role in the transition of transgender women, as typically masculine facial features can make it difficult to be perceived as the correct gender. One of the difficulties with FFS is that the indications and results are very subjective and difficult to quantify. The research presented in this thesis focuses on the use of facial three-dimensional surface scans to analyse the facial morphology of transgender women. This can lead to a better understanding and quantification of the indications and results of FFS.

The facial morphology of transgender women is analysed and compared to the facial morphology of cisgender male and female faces. This is done by performing a principal component analysis (PCA) on the coordinates of facial three-dimensional surface scans. This is done for the entire facial shape, as well as for the forehead, nose, chin and zygoma region separately. The effect of different FFS procedures on the facial shape is evaluated by comparing preoperative and postoperative facial 3D surface scans of transgender women. Furthermore, the output of the PCA is used to propose a scoring system that can be used to describe the masculinity or femininity of the shape of the face in one single value. Lastly, an artificial intelligence network is trained to be able to classify facial 3D surface scans as either male or female.

The comparison between cisgender male, cisgender female and transgender female faces based on the principal component scores did offer some insight into the facial features that differ between cisgender and transgender faces. However, analysis of the individual principal components was demonstrated to be of limited value in comparing preoperative and postoperative scans of transgender women. Combining the individual principal component scores into one single measure of the masculinity and femininity of the face showed promising results for an objective evaluation of the indications and results of FFS. Additionally, classification of the faces of transgender women using an artificial intelligence model illustrated both the need for FFS, as well as the successful results of FFS in terms of creating a feminine shape of the face.

The study presented in this thesis demonstrated different approaches to evaluate the facial shape of transgender women before and after FFS. The possible added value of 3D surface scans and facial shape analysis has been illustrated. However, further improvements of the proposed methods are needed before clinical implementation can be considered.

Keywords: 3D shape analysis, Facial feminization surgery, Facial morphology, Sexual dimorphism, Transgender

Abbreviations

+3SD	Three standard deviations from the mean in the positive direction
-3SD	Three standard deviations from the mean in the negative direction
2D	Two-dimensional
3D	Three-dimensional
ANOVA	Analysis of Variance
AI	Artificial Intelligence
CI	Confidence Interval
Cis.	Cisgender
FFNN	Feed-forward Neural Network
FFS	Facial Feminization Surgery
GPA	Generalised Procrustes Analysis
LOO	Leave-one-out
MANOVA	Multivariate Analysis of Variance
PC	Principal Component
PCA	Principal Component Analysis
ReLu	Rectified Linear unit
SD	Standard Deviation
SSM	Statistical Shape Model
Trans.	Transgender

Table of contents

Abstract **i**

Abbreviations **ii**

General introduction **1**
Gender dysphoria 1
Facial feminization surgery 1
Scope of the thesis 2

1 Principal component analysis of facial 3D surface scans; Analysing the facial morphology of transgender women **3**
1.1 Introduction 3
1.2 Methods 4
1.2.1 Dataset 4
1.2.2 Preprocessing 5
1.2.3 Principal component analysis 5
1.2.4 Statistical analysis 6
1.3 Results 6
1.3.1 Principal component analysis 6
1.3.2 Statistical analysis 8
1.4 Discussion 9
1.4.1 Interpretation of results 9
1.4.2 Study limitations 11
1.5 Conclusion 13

2 Principal component analysis of specific facial regions; Analysing the facial morphology of transgender women **14**
2.1 Introduction 14
2.2 Methods 15
2.3 Results 16
2.3.1 Results for the forehead region 16
2.3.2 Results for the nose region 17
2.3.3 Results for the chin region 19
2.3.4 Results for the zygoma region 21
2.4 Discussion 23
2.4.1 Interpretation of the results 23
2.4.2 Study limitations 24
2.5 Conclusion 25

3 The effect of facial feminization surgery on the facial shape; Comparing preoperative and postoperative 3D surface scans of transgender women 26
3.1 Introduction 26
3.2 Methods 27
3.2.1 Dataset 27
3.2.2 Preprocessing 27
3.2.3 Principal component scores and Z-scores 27
3.2.4 Visualization of the average change in facial shape 28
3.3 Results 28
3.4 Discussion 28
3.4.1 Interpretation of the results 29
3.4.2 Study limitations 30
3.5 Conclusion 31

4 Scoring the masculinity or femininity of the face of transgender women before and after facial feminization surgery based on principal component analysis 32
4.1 Introduction 32
4.2 Methods 33
4.2.1 Dataset 33
4.2.2 Calculating the masculinity/femininity score 33
4.2.3 Scoring the faces of cisgender men and women 34
4.2.4 Scoring the faces of transgender women 35
4.3 Results 35
4.3.1 Masculinity/femininity scores of cisgender men and women 35
4.3.2 Masculinity/femininity scores of transgender women 36
4.4 Discussion 37
4.4.1 Interpretation of results 37
4.4.2 Study limitations 38
4.5 Conclusion 39

5 Gender classification of the faces of transgender women before and after facial feminization surgery using a feed-forward neural network. 40
5.1 Introduction 40
5.2 Methods 41
5.2.1 Dataset 41
5.2.2 The feed-forward neural network 41
5.2.3 Evaluation of the performance of the model 42
5.2.4 Classification of transgender faces 42
5.3 Results 42
5.4 Discussion 43
5.4.1 Interpretation of the results 43
5.4.2 Study limitations 44
5.5 Conclusion 44

General discussion & conclusion 46

References **48**

Appendix **51**
A. Results of the PCA of the face 51
B. Results of the PCA for separate facial regions 56
C. Results of the PCA for the forehead region 57
D. Results of the PCA for the nose region 59
E. Results of the PCA for the chin region 61
F. Results of the PCA for the zygoma region 63
G. Results of the 5-fold cross-validation of the classification network for the full face. 65

General introduction

GENDER DYSPHORIA

A transgender individual is someone whose gender identity differs from the gender they are assigned at birth. The number of transgender individuals has strongly increased in the past couple of decades.[1,2] In the Netherlands 1.1% of individuals assigned male at birth and 0.8% of individuals assigned female at birth identify stronger with the other gender than with the gender assigned at birth.[3] The distress resulting from the conflict between assigned gender and gender identity is defined as gender dysphoria.[4] There are different therapies available that aid in transitioning from male to female, and vice versa. These therapies include psychotherapy, hormonal treatments and gender-affirmative surgery.[5,6] About 0.6% of individuals assigned male at birth, and 0.2% of individuals assigned female at birth wish to obtain hormone therapy or gender-affirmative surgery to reduce gender dysphoria.[3]

Gender-affirmative surgery aims to reduce gender dysphoria by surgically modifying a person's external appearance to match their gender identity.[7] There is a large variety of gender-affirmative surgeries to achieve masculinization or feminization of the body. However, not all patients undergo the same set of procedures, as this depends on individual wishes and causes of gender dysphoria. For individuals who are transitioning from female to male the most common surgeries are mastectomy (removal of the breast tissue), hysterectomy (removal of the uterus), metoidioplasty (lengthening of the hypertrophied clitoris), phalloplasty (construction of a penis) and facial masculinization surgery.[8,9,10] For individuals who are transitioning from male to female frequently performed surgeries are orchiectomy (removal of the testicles), vaginoplasty (construction of a neovagina), mammaplasty (breast augmentation), thyroid chondroplasty (reduction of thyroid cartilage), chondrolaryngoplasty (vocal cord surgery) and facial feminization surgery.[8,9,11] The research presented in this thesis focuses on this last group of surgeries, the facial feminization surgery.

FACIAL FEMINIZATION SURGERY

Facial feminization surgery (FFS) comprises a broad range of craniomaxillofacial surgical procedures with the objective to change masculine facial features into feminine features.[12,13] As the face is one of the most visible determinants of gender, gender-affirmative facial surgery can play an important role in the transition of transgender individuals.[14] Especially for transgender women, typically masculine facial features can make it difficult to be perceived as the correct gender.[14] In general, a masculine face is more angulated with a pronounced jaw and chin, while a feminine face is more rounded and soft. Women typically have a shorter forehead, no supraorbital bossing, a smaller nose, more pronounced zygomatic prominences, fuller lips, a smaller mandibular width, and a more tapered chin.[15] Figure 1 shows an example of a male and female face, in which the differences between a typically masculine and a typically feminine face can be observed. Because different aspects of the face play a role in creating an overall feminine or masculine appearance, a wide variety of surgical procedures can be performed to create more feminine looking face. For example, FFS can include forehead contouring, a brow lift, correction of the hairline by scalp advancement, rhinoplasty, genioplasty, zygoma osteotomy, bimaxillary osteotomy, mandibular angle reduction and a thyroid shave.[12,13,16]

FFS has proven to be beneficial to the mental health and quality of life of transgender women, as well as improve their appearance and satisfaction.[5,14] Although the demand for FFS keeps increasing, most training programs have minimal exposure to FFS.[17] As FFS is still a relatively new and developing surgical discipline, there is a need for evidence-based guidelines and improvement of the surgical procedures.[17,18] One of the difficulties with FFS is that the indications are very subjective and difficult to quantify. This results in a problem with insurance coverage, as FFS is often considered as a cosmetic procedure instead of a medical necessity.[19] Furthermore, it is difficult for the patient to form realistic expectations for FFS, due to the limited ability to accurately predict and simulate the surgical outcome.[20,21] Finally, it remains difficult to quantify the results of FFS, making it difficult to evaluate the outcome and effectiveness of the surgery.[7,22]

FIGURE 1. Frontal and profile view of a typical male and female face, illustrating the difference between male and female facial morphology.[12]

SCOPE OF THE THESIS

Based on the highlighted need for further development in the field of FFS, the research presented in this thesis focuses on the use of facial three-dimensional (3D) surface scans to analyse the faces of transgender women. 3D surface scans can be used to capture the shape of the face, to allow for an objective and quantitative analysis of facial morphology. This can lead to a better understanding and quantification of the surgical indications and results of FFS, as well as offer the ability to create a simulation of the desired surgical outcome for transgender women.

The research can be divided into several different parts. The first part investigates the facial shape of transgender women prior to FFS, in comparison to cisgender male and female faces. These comparisons are based on a Principal Component Analysis (PCA) of the face (Chapter 1). Following the same methods, separate facial analyses are performed for the forehead, nose, chin and zygoma, allowing for more specific evaluations of these facial regions (Chapter 2). Next, the effect of different FFS procedures on the facial shape is assessed by comparing preoperative and postoperative 3D surface scans of transgender women (Chapter 3). The masculinity or femininity of the face and the specific facial regions is captured in one single score, using the results of the PCA (Chapter 4). Furthermore, the outcome of the PCA is used to train an artificial intelligence network to be able to classify facial 3D surface scans as male of female. This network is then applied to preoperative and postoperative 3D surface scans of transgender women to evaluate how these scans are classified (Chapter 5). The thesis is concluded with a general discussion about the added value of 3D surface scans and facial analysis within the field of FFS.

Chapter 1. Principal component analysis of facial 3D surface scans; Analysing the facial morphology of transgender women

1.1. INTRODUCTION

The face is one of the main determinants we use to recognize other people and plays an important role in communication, social interactions and expression of emotions. There is a large variation in facial morphology between individuals, established by both environmental and genetic factors. Facial morphology can offer a lot of information about an individuals health or demographics, such as age, gender and race. Therefore, facial shape analysis has become an important tool in various fields, such as medicine, biometrics, anthropology and forensics.

Traditional methods to analyse facial shape are based on two dimensional (2D) images or measurements. Information about the shape of the face is often described by linear or angular measurements between anatomical landmarks. This offers limited information about the complex 3D shape of the face. With the developments in 3D imaging technologies, more advanced methods to describe the facial shape have been established, including geometric morphometrics. In geometric morphometrics the facial shape is not described using measurements between individual landmarks, but rather by the geometric configuration of the complete set of anatomical landmarks.[23] This allows for a more comprehensive analysis of the complex facial shape.

To analyse the facial shape variation, Statistical Shape Models (SSMs) can be used. SSMs are mathematical models that describe the variation in shape within a dataset by defining the average shape and possible deviations from this average.[24] SSMs are constructed by performing a Principal Component Analysis (PCA). PCA breaks down the data into its mean and the covariance matrix that describes the variation around that mean. This reduces the dimensionality in the dataset while retaining most of the variance. By further decomposing the covariance matrix, PCA linearly transforms the original data into a new set of uncorrelated variables, the principal components (PCs). The PCs are ordered based on how much of the variance in the dataset is explained by each PC. For the PCA in the context of facial shape analysis this means that each PC represents a different way in which a face can vary from the average facial shape. Any possible facial shape can be described by a weighted combinations of the PCs. The facial shape variation within a dataset can thus be represented as an SSM, describing the mean facial shape and a collection of PC's that explain the possible variance around this mean. The SSM offers a compact representation of the facial shape variation that can be used to analyse, compare and transform facial shapes.

Within this research, SSMs are constructed with a PCA-approach to analyse the facial shape of transgender women in the context of sexual dimorphism. Sexual dimorphism refers to the physical differences between males and females of the same species. These differences are also present in facial morphology, as males tend to have a more angulated facial shape as opposed to the more rounded facial shape of females. Furthermore, there are differences in shape and size of specific facial features such as the brow ridge, nose, chin and jaw.[15] Previous studies showed a significant difference between male and female faces, both in 2D measurements[25,26] and 3D analyses[27,28,29,30]. However, there is limited research available about how the faces of transgender women compare to the faces of cisgender male and female faces. A lot of the typically male facial features can be seen in the faces of transgender women, explaining the wish for facial feminization surgery. However, facial morphology is also influenced by the gender-affirmative hormones that are part of the treatment of a lot of transgender women. This can alter the facial shape, introducing more typically feminine features such as fuller cheeks and a smaller jaw.[31]

The aim of the research presented in this chapter is to analyse the facial morphology of transgender women prior to FFS, in comparison to cisgender male and female faces. The facial morphology is captured using 3D surface scans, followed by geometric facial shape analysis using a PCA-approach. Statistical analysis is conducted to determine which PCs illustrate a difference between cisgender male, cisgender female and transgender female faces. This can offer more insight into the sexual dimorphism of the face, as well as how the faces of transgender women compare to cisgender male and female faces.

1.2. METHODS

1.2.1 Dataset

The data used in this study comprises of two separate datasets, one dataset containing 3D surface scans of cisgender male and female faces and one dataset containing 3D surface scans of the faces of transgender women. For the cisgender dataset, 782 3D surface scans are obtained from the Headspace dataset, created by the Alder Hey Craniofacial Unit (Liverpool, UK) and the Department of Computer Science of the University of York (York, UK).[32] Additionally, 454 scans are added from an internal database of the Department of Oral and Maxillofacial Surgery of the Radboudumc (Nijmegen, the Netherlands). This results in 1236 scans to be included in the cisgender dataset (58.09% female, 41.91% male). All scans were obtained with the 3DMDhead scanning system (3dMD, Atlanta, USA).

The transgender dataset comprises of 203 facial 3D surface scans of transgender women, who are receiving gender-affirmative treatment at the Amsterdam UMC (Amsterdam, The Netherlands). The scans are obtained using the Artec Leo handheld 3D scanner (Artec 3D, Luxembourg, Luxembourg) and converted to 3D meshes using the Artec Studio Professional software (Artec 3D, Luxembourg, Luxembourg). The transgender women in this dataset have not yet undergone FFS. For both the cisgender and transgender dataset only Caucasian subjects aged 18-50 years without known genetic disorders or craniofacial dysmorphism are included. Further exclusion was based on facial expression, insufficient quality of the scan, or the absence of information such as gender and age.

1.2.2 Preprocessing

To allow for a more standardised analysis of the facial scans, each 3D surface scan in the datasets is represented as a homologous 3D mesh using the MeshMonk algorithm.[33] MeshMonk is an open-source surface registration toolbox, that allows for a semi-automatic placement of a template onto a facial 3D surface scan.[33,34] The generic template consists of 7160 vertices, which are placed at specific anatomical positions. Meaning each face is represented as a standard set of points, in which each anatomical position corresponds to the same vertex in all scans. Five anatomical landmarks (right and left endocanthion, pronasale and left and right cheilion) were manually located on each 3D scan as well as on the generic template. These landmarks were used to create an initial fit of the generic template onto the scan. The template is then fitted onto the 3D surface scan using scaled rigid registration followed by non-rigid registration. Finally, all scans are aligned using Generalised Procrustes Analysis (GPA), to ensure corresponding location, orientation and scale. Figure 2 shows an example of the facial template before and after being fitted to a 3D surface scan.

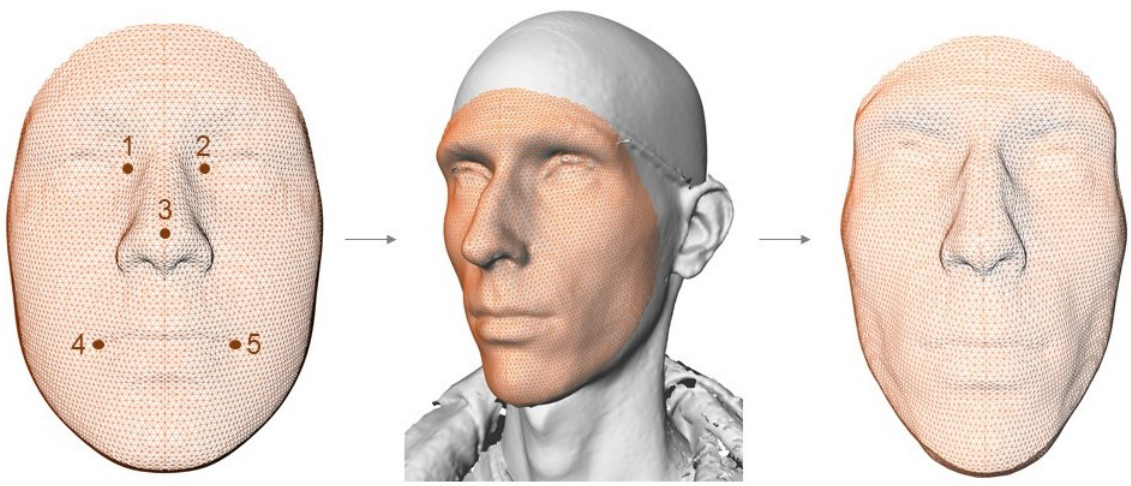

FIGURE 2. Example of the generic facial template being fitted to a 3D surface scan. The five anatomical landmarks are annotated on the generic template, 1: Right endocanthion, 2: Left endocanthion, 3: Pronasale, 4: Right cheilion, 5: Left cheilion.

1.2.3 Principal component analysis

After preprocessing, the initial PCA was performed on the x-, y- and z-coordinates of the vertices of the 1236 facial 3D surface scans included in the cisgender dataset. This results in the PC scores of the individual faces in the cisgender dataset. The PC scores of the faces in the transgender dataset are calculated using the eigenvectors of the covariance matrix resulting from the initial PCA. The PCs that cumulatively account for 80% of the total variance in the cisgender dataset are used in further visualisations and statistical analysis. To visualize the shape variability captured in each individual PC, faces are constructed for which the PC score for that specific PC differs three standard deviations from the mean face, both in positive (+3SD) and negative (-3SD) direction. Distance maps are created illustrating the differences between the -3SD and +3SD reconstructed faces in the direction perpendicular to the facial surface. From these visualisations the anatomical meaning of each PC is evaluated. The preprocessing of the scans, the PCA and the visualizations are performed using MATLAB R2022a (The MathWorks Inc., Natick, USA).

1.2.4 Statistical analysis

The PC scores are divided into three groups: cisgender male, cisgender female and transgender female. To identify the PCs that illustrate a difference between these three groups, statistical analysis is performed. For each PC, the PC scores of the cisgender male faces, cisgender female faces and transgender female faces are compared using a one-way multivariate analysis of variance (MANOVA) (CI=95%). If the analysis shows a significant difference among the three groups, post hoc analysis is performed to identify which specific groups show a statistically significant difference. Statistical analysis is performed using IBM SPSS Statistics 27 (IBM Corp., Armonk, USA).

1.3. RESULTS

1.3.1 Principal component analysis

The first twelve PCs cumulatively account for 80% of the variability of the facial shape within the dataset. An overview of the percentage of variability explained by the individual PCs can be found in Appendix A.1. For each PC, the shape variability is visualized as the -3SD reconstruction, the +3SD reconstruction and the distance map illustrating the difference between these two reconstructions. Figure 3 shows an example of the visualizations for PC2. A comprehensive overview of the visualizations of the first 12 PCs can be found in Appendix A.2. The distance maps for the first 12 PCs are shown in figure 4. From the visualisations the anatomical meaning of each PC is evaluated, as described in table 1.

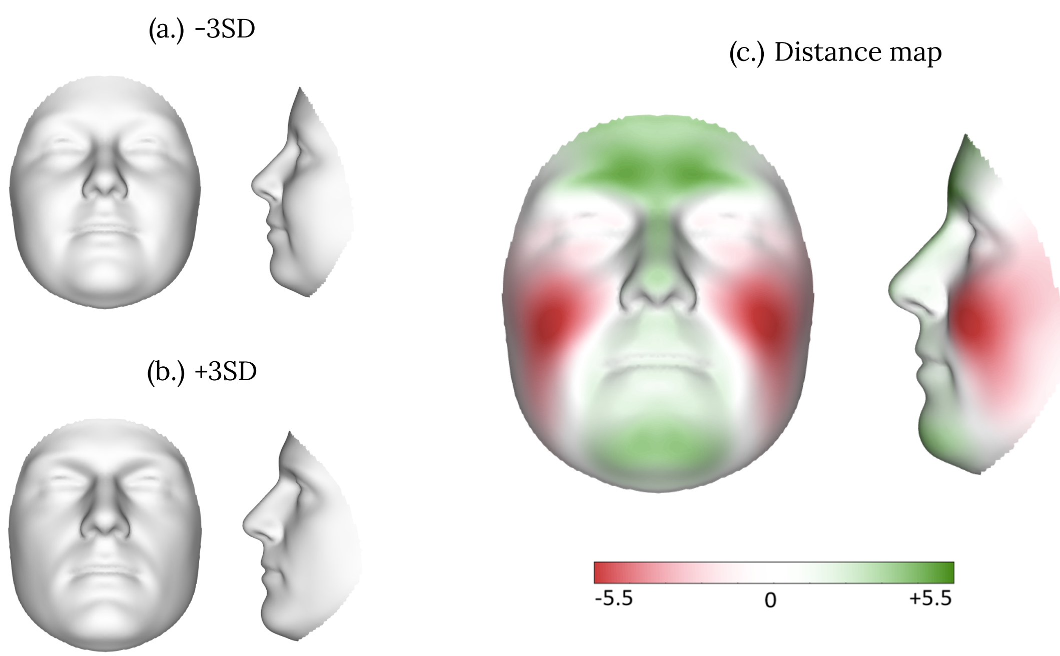

FIGURE 3. Visualization of the shape variability captured in principal component 2. (a.) Reconstructed face that differs three standard deviations from the mean face in the negative direction (-3SD). (b.) Reconstructed face that differs three standard deviations from the mean face in the positive direction (+3SD). (c.) Distance map illustrating the differences between the -3SD and +3SD reconstructed faces. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

TABLE 1. The anatomical meaning of the shape variability within the face captured in the first 12 principal components.

PC	Main attributes in facial shape variability captured in each principal component
PC1	Width of the face, with higher PC scores corresponding to a broader face.
PC2	Changes in the forehead, cheeks, nose and chin. Higher PC scores result in a more prominent brow ridge, a larger nose, more projection of the chin and less volume in the cheeks.
PC3	Position of the nose, mouth and chin within the face. The higher the PC score, the lower the nose, mouth and chin are positioned within the face, also resulting in a increased length of the nose.
PC4	Projection of the upper jaw in comparison to the chin. High PC scores correspond to less anterior projection of the upper jaw and more anterior projection of the chin.
PC5	Position of the eyes relative to the infraorbital region. As the PC scores increase, the projection of the infraorbital region decreases and the eyes become more protruding.
PC6	Forward projection of the nose, with higher PC scores representing a larger, more prominent nose.
PC7	Shape of the superior part of the forehead, with higher PC scores resulting in a flatter forehead and lower PC scores in a more sloped forehead.
PC8	Volume of the lower cheeks and the zygomatic projection. High PC scores are associated with less zygomatic projection and an increase in volume in the lower cheeks.
PC9	Transition from the chin to the neck, with higher PC scores representing a more steep transition and lower PC scores representing a more gradual transition with more volume below the chin.
PC10	Shape of the jaw, with high PC scores representing a more squared jaw.
PC11	Change in shape around the mouth, where the lips protrude more with increasing PC scores.
PC12	Asymmetric facial shape variation, where a high PC score represents a counterclockwise rotation.

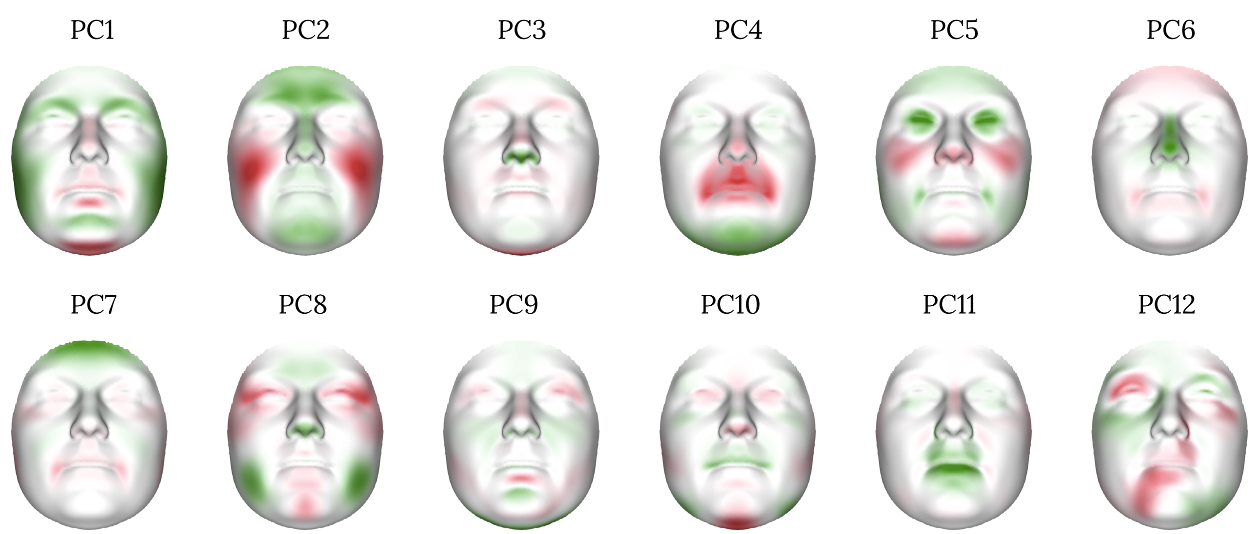

FIGURE 4. Distance maps illustrating for each of the principal components (PCs) the differences between the reconstructed faces that differ three standard deviations from the average face in the positive and in the negative direction. The distances are measured in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

1.3.2 Statistical analysis

Statistical analysis was performed to compare the PC scores of the faces of cisgender males, cisgender females and transgender females. The descriptive statistics and distribution of the PC scores for the different groups can be found in Appendix A.3. A one-way MANOVA (CI=95%) is performed using the first twelve PCs, showing a significant difference in PC scores among the three groups, $F(24, 2840) = 59.781$, $p<.001$; Pillai's Trace = 0.671, partial $\eta 2 = 0.34$. For each of the twelve PCs a follow-up ANOVA was performed, showing a statistically significant difference among the three groups (cisgender male, cisgender female and transgender female) for PCs 2, 3, 4, 7, 8, 9 and 10. No statistical difference was found among the groups for PCs 1, 5, 6, 11 and 12. PC4 and PC10 are excluded from further analysis due to the low effect size (partial $\eta 2$), respectively .020 and .010. The results are shown in table 2.

TABLE 2. Results of the ANOVA for the first twelve principal components.

PC	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial $\eta 2$
PC1	88179.239	2	44089.619	2.715	.067	.004
PC2	5678751.593	2	2839375.797	403.082	.000	.361
PC3	623155.908	2	311577.954	69.763	.000	.089
PC4	125259.179	2	62629.590	14.461	.000	.020
PC5	14092.742	2	7046.371	2.751	.064	.004
PC6	6407.175	2	3203.588	1.327	.266	.002
PC7	97410.169	2	48705.084	29.280	.000	.039
PC8	143553.580	2	71776.790	49.942	.000	.065
PC9	113785.153	2	56892.576	41.832	.000	.055
PC10	16214.035	2	8107.017	7.176	.001	.010
PC11	3890.342	2	1945.171	1.873	.154	.003
PC12	2759.967	2	1379.983	1.357	.258	.002

For the PCs that showed a statistical difference among the three groups together with a sufficient effect size (PCs 2, 3, 7, 8 and 9), post hoc analysis is performed using the Scheffe method. This offers individual comparisons between the groups of cisgender male, cisgender female and transgender female faces. The results are shown in table 3. A statistically significant difference ($p < .05$) between the groups of cisgender male and cisgender female faces is found in all PCs, except for PC4 ($p=0.833$). Transgender females compared to cisgender males show a statistically significant difference for all PCs except for PC3 ($p=.154$) and PC7 ($p=.963$). Finally, when comparing transgender females to the cisgender females, statistically significant differences are observed in all PCs except for PC8 ($p=.393$) and PC10 ($p=.948$).

TABLE 3. Results of the post hoc analysis (Scheffe's method) to compare the PC scores of the face for cisgender males, cisgender females and transgender females.

					95% Confidence Interval	
PC	Comparison between groups	Mean Difference	Std. Error	Sig.	Lower Bound	Upper Bound
PC2	Cis. Male - Cis. Female	132.427*	4.848	.000	120.549	144.306
	Trans. Female - Cis. Male	-27.621*	6.980	.000	-44.724	-10.517
	Trans. Female - Cis. Female	104.807*	6.699	.000	88.393	121.220
PC3	Cis. Male - Cis. Female	-38.126*	3.860	.000	-47.585	-28.667
	Trans. Female - Cis. Male	-10.750	5.558	.154	-24.369	2.870
	Trans. Female - Cis. Female	-48.876*	5.334	.000	-61.945	-35.806
PC7	Cis. Male - Cis. Female	-16.218*	2.356	.000	-21.990	-10.445
	Trans. Female - Cis. Male	-.931	3.392	.963	-9.243	7.380
	Trans. Female - Cis. Female	-17.149*	3.255	.000	-25.125	-9.173
PC8	Cis. Male - Cis. Female	21.569*	2.190	.000	16.204	26.935
	Trans. Female - Cis. Male	-17.432*	3.153	.000	-25.158	-9.706
	Trans. Female - Cis. Female	4.137	3.026	.393	-3.277	11.551
PC9	Cis. Male - Cis. Female	-9.676*	2.130	.000	-14.896	-4.457
	Trans. Female - Cis. Male	27.904*	3.067	.000	20.388	35.419
	Trans. Female - Cis. Female	18.227*	2.943	.000	11.0154	25.439

* The mean difference is significant at the .05 level.

1.4. DISCUSSION

The aim of the research presented in this chapter was to analyse the facial morphology of transgender women, in comparison to cisgender male and female faces. The PC scores of the 3D surface scans of cisgender male, cisgender female and transgender female faces are calculated and compared using statistical analysis.

1.4.1 Interpretation of results

Five of the twelve PCs showed a statistically significant difference among the three groups, as well as a sufficient effect size. For all 5 PCs a statistically significant difference was found between the groups of cisgender men and cisgender women, meaning all 5 PCs illustrate 'features' that play a role in the sexual dimorphism of the face. For PC2 this was expected, as this PC explains changes in shape in the forehead, cheeks, nose and chin that correspond to existing literature about sexual dimorphism.[15] The PC scores of cisgender men are observed to be higher than the scores of cisgender females, meaning cisgender men have a more prominent brow ridge, a larger nose, more projection of the chin and less volume in the cheeks. PC2 also showed by far the largest effect size (partial $\eta 2$ = .361) for the differences among the three groups, meaning these differences are most likely to be of practical or clinical value. For the other PCs the effect size is a lot smaller, and the presence of sexual dimorphism is more difficult to explain as these anatomical meanings do not directly correspond to features of sexual dimorphism as described in literature. As the overall facial shape or shape of specific facial regions does not depend on one single PC but rather on the combined scores of all PCs, it is possible to observe sexual dimorphism in PCs that is difficult to explain when looking at the individual meanings of the PCs.

When comparing the faces of transgender women with the faces of cisgender women, a statistically significant difference is found in four of the five PCs. This means that for almost all 'features' that differ between cisgender male and cisgender female faces, this difference is also present between transgender women and cisgender women. This is in accordance with the expectations that the faces of transgender women do not show a lot of the typically feminine facial features. This also explains the wish for feminization of the face for a lot of transgender women. However, there are also statistically significant differences observed between the faces of transgender women and cisgender men, indicating that the facial shape of transgender women does not completely correlate with the facial shape of cisgender men. Differences between the faces of transgender women and cisgender men can be the result of gender-affirmative hormone treatment or previous treatment with puberty blockers. Gender-affirmative hormone treatment can change the facial shape of transgender women, introducing more typically feminine features such as fuller cheeks and a smaller jaw.[31] As the volume of the cheeks is part of the facial shape variation captured in PC2, this might explain the difference in PC scores between transgender women and cisgender men for PC2. However, with the limited availability of comprehensive studies about the influence of gender-affirmative hormone treatment on the facial shape, it is difficult to corroborate these findings. Another possible source for the differences in facial shape between transgender females and cisgender males can also be attributed to the possible presence of dermal fillers or surgical procedures that have previously been performed, which might have been missed during the selection of individuals for the dataset.

For some of the PCs a significant difference between the cisgender male and female faces would have been expected based on the anatomical meaning of the PC, but could not be observed in this study. For example, both the size of the nose and the shape of the jaw have been described in literature as features that show sexual dimorphism in the face.[15] However, no statistically significant difference was found for the forward projection of the nose (PC6) or the shape of the jaw (PC10). The shape of the nose is also influenced by the PC scores of PC2 and PC3, while the shape of the jaw is also influenced by PC2 and PC8. These PCs did show a statistically significant difference between cisgender male and female faces. It is therefore possible that the main differences between male and female faces in terms of the shape of the nose and the jaw is mostly captured in different PCs. Furthermore, there is no difference observed among the three groups for PC11, which represents the shape around the mouth. Some previous studies show a difference in shape of the mouth and lips between male and female faces.[25,35,36] Apart from the influence that other PCs can have on the shape of the mouth, variation in this area is also expected to be influenced by slight variations in facial expression. This could contribute to the lack of sexual dimorphism observed within this region. In the end, the overall shape of the face and facial regions is determined by the combination of all PCs together, so not all facial features can be attributed to one single PC.

As facial morphology is influenced by a lot of different factors, not all facial shape variation can be attributed to differences in gender. Factors such as age, length, weight and race might be responsible for part of the facial shape variation described by the PCs. Therefore, it is in accordance with the expectations that not all PCs show a significant difference between the three groups. For example, PC1 explains the variation in width of the face, which showed no statistically significant difference between the three groups. This means that although the width of the face explains a large part of the facial shape variability, there is no difference observed between cisgender male, cisgender female and transgender female faces. This corresponds with existing literature showing no sexual dimorphism for the width-to-height ratio of the face.[37,38] In general, the size of the male face is considered to be larger than the female face.[35] However, due to the scaling of the 3D surface scans in the this study, differences in general size of the face could not be investigated. For the remaining PCs that showed no statistically significant differences or a sufficient effect size, this is in accordance with expectations as these 'features' have not been reported in literature to be associated with sexual dimorphism of the face. It was therefore expected that these PCs show no significant difference between the faces of cisgender men, cisgender women and transgender women.

1.4.2 Study limitations

Some of the study limitations have already been briefly discussed above. Most importantly, the limited specificity of the anatomical meanings of the PCs. Although in this study the meaning of the individual PCs is determined and used to assess the facial shape difference, it should be considered that the overall shape of the face is determined by the combination of all PCs. Meaning, multiple PCs can affect the same facial features or regions. Additionally, one single PC can influence multiple facial features at the same time. It is therefore not possible to attribute one single facial feature to one single PC. However, adding anatomical meanings to the PCs provides more insight into the facial regions and structures that that display the most facial shape variance. This also makes it possible to analyse differences between male and female faces, as well as analyse the facial shape of transgender women. As individual PCs often capture facial shape variability that include multiple parts of the face, it is difficult to extract information about the shape of specific facial features or facial regions. Within the context of FFS it would be useful to be able to analyse specific facial regions independently. A separate PCA for different facial regions could offer more specific information about the facial shape of different regions and therefore make the analyses more applicable in the field of FFS. These analyses for specific facial regions will be performed and described in Chapter 2.

The statistical analyses in this study are performed using the first 12 PCs, as these cumulatively explain 80% of the variance within the dataset. It can be argued whether 80% explained variability is sufficient to explain the facial shape, or that a larger threshold should be used. By limiting the number of PCs to a specific threshold of explained variability, noise and redundant information is removed and will therefore not influence the results of the analysis. Meaning, specific individual deviations in facial shape or local inaccuracies in the 3D meshes will not influence the facial analysis. Using a higher threshold, and thus more PCs, can lead to a more comprehensive analysis with a more accurate representation of the full facial shape. However, with each additional PC explaining only a very small percentage of the variability within the dataset, it should be questioned whether analysis of these individual PCs holds any clinical relevance.

Another study limitation is the generalizability of the results of the study. Although facial features can vary across European populations[39], the used cisgender dataset offers a fairly reliable representation of the Caucasian male and female face. In this study the facial shape variability is only analysed for Caucasian individuals. Because both facial morphology and sexual dimorphism in the face are influenced by race[40], the results of this study are not directly applicable to the faces of non-Caucasian individuals. As the aim of the facial shape analysis is to illustrate differences between cisgender male, cisgender female and transgender female faces, it is preferred to limit the facial shape variability introduced by other factors such as race. Ultimately, the methods proposed in this chapter can be repeated for different races, provided there are sufficient 3D surface scans available for both the cisgender and transgender dataset. Expanding the research to include different races is especially important within the context of FFS, as a large part of the patient population is non-Caucasian.

For the dataset of transgender women there are some more factors to take into consideration, most importantly the selection bias within this dataset. The 3D surface scans of transgender women are collected during preoperative consultations concerning facial feminization surgery. This means that all transgender women included in this dataset wish to undergo FFS. This may lead to a higher prevalence of faces with more masculine features within the dataset, as transgender women with less masculine facial features are less likely to require FFS. The used dataset may therefore not be reflective for the entire population of transgender women. However, as this study focuses on the facial shape differences in respect to facial feminization surgery it can be justified to select only patients who are planning to undergo these procedures. As mentioned above, the facial shape of transgender women is also influenced by the gender-affirmative hormone therapy that is part of the treatment for a lot of transgender women. The individuals that are included in the transgender dataset have had gender-affirmative hormone therapy for at least one year, as this is one of the requirements to be eligible for FFS. However, there are a lot of factors associated with gender-affirmative hormone therapy that can influence the facial shape variability. Such factors include differences in gender-affirmative hormones, duration of the hormone treatment, age at which the hormone treatment was started and prior treatment with puberty blockers. The exact way in which these factors influence the facial shape variability and sexual dimorphism of the face is unknown and would make an interesting topic for future research.

A possible source for inaccuracies lies within the preprocessing of the 3D surfaces scans that are used to analyse the facial shape. 3D surface scans have become a popular imaging modality in craniofacial practice, due to the possibility to capture highly accurate surface scans without the use of ionizing radiation.[41] The 3D surface scans used in this study are captured by two different camera systems. The scans in the cisgender dataset are captured using the 3DMDhead scanning system (3dMD, Atlanta, USA) and the scans in the transgender dataset are captured using the Artec Leo handheld 3D scanner (Artec 3D, Luxembourg, Luxembourg). Both the 3dMDhead system and the Artec Leo 3D scanner are reported to have a clinically sufficient accuracy of respectively 0.2 mm and 0.1 mm.[42] However, the quality of the scan also depends on factors such as lighting, camera calibration, movement of the patient, scan protocol and postprocessing of the scans.[43] 3D surface scans that visually showed insufficient quality or large inaccuracies have been excluded from the dataset. The 3D surface scans are represented as homologous 3D meshes with corresponding vertices. The MeshMonk registration is reported to have an average error of 1.26 mm.[33]. This error is determined across 19 landmarks placed within the central region of the face. The accuracy of the registration is expected to be lower around the edges of the mesh, as visual inspection of the results of the registration in this study shows larger variations in the outline of the mesh.

Another limitation of using the MeshMonk algorithm with the corresponding template, is that the template does not cover the entire facial surface. The coverage of the facial template also varies as a result of facial shape differences. As can be seen in figure 2, the template only covers the lower part of the forehead. Especially in individuals with a more prominent forehead and brow ridges, the template does not extend to the full height of the forehead. The same applies to the mandible, as the template does not extend around the edges of the mandible. Shape variation in the mandibular angle could therefore not be analysed within this research. This is an important limitation, as the mandibular angle is considered to be an important distinguishing feature between male and female faces.[44,45] Reduction of the mandibular angle is therefore one of the possible procedures that can be performed in FFS. For future research it would be recommended to develop a facial template that covers a larger part of the forehead and mandible.

1.5. CONCLUSION

This chapter illustrated the use of a PCA-approach for geometrical shape analysis of the facial morphology of cisgender men, cisgender women and transgender women. The facial shape variability captured in the first twelve PCs is visualized and described, and the PC scores are compared using statistical analysis. Although it is difficult to attribute anatomical meanings to individual PCs, it does give more insight into the facial shape differences between cisgender males, cisgender females and transgender females. The differences between male and female faces found in this study largely correlate to the sexual dimorphism in the face described in literature. Most important distinguishing features include the prominence of the brow ridge, size of the nose, projection of the chin and volume of the cheeks. When analysing the facial shape of transgender women, it could be observed that there is a statistically significant difference between the facial shape of transgender women and cisgender women for most of the PCs that illustrate sexual dimorphism. This confirms the need for facial feminization surgery. However, there are also some differences observed between the transgender female faces and cisgender male faces, possibly indicating the effects of gender-affirmative hormone treatment. Further research into the effects of gender-affirmative hormone treatment is needed to provide more insight into these differences. A separate PCA for different facial regions could offer more specific information about facial features, making the analyses more applicable in the field of facial feminization surgery.

Chapter 2. Principal component analysis of specific facial regions; Analysing the facial morphology of transgender women

2.1. INTRODUCTION

The previous chapter described the possibility of analysing the facial morphology of transgender women using a PCA of the entire facial shape. The PCA was performed using the vertex coordinates of the faces of cisgender men and women. The PC scores of the faces of transgender women were later calculated using the covariance matrix of the initial PCA. The PC scores of the twelve PCs that explained the most facial shape variability were then compared for transgender women, cisgender men and cisgender women. A problem with this approach is that a lot of the PCs represent facial shape variation across a large part of the face. It is therefore difficult to describe or quantify the facial shape of specific regions using this method. A possible solution would be to look at different facial regions separately, instead of analyzing the entire face at once.

Especially in the context of FFS, it is important to be able to assess different facial regions independently. FFS comprises of several different craniomaxillofacial surgical procedures, which can be performed separately or in an arbitrary combination of different procedures. Which specific procedures are performed depends on the facial shape as well as the wishes of the patient. Although the choice of procedures should always be based on the source of the dysphoric feelings of the patient, not all patients have a clear vision of which facial parts cause the most distress. For this purpose it would be beneficial to be able to assess the facial shape as well as the shape of specific facial regions. This would allow for a more specific assessment of which facial regions display more masculine features, and which facial regions might already present quite feminine.

The facial regions that play an important role in relation to FFS are the forehead, nose, zygoma, chin and jaw, as these are areas that are associated with sexual dimorphism of the face.[40,30] Typically masculine facial features can also be present in the faces of transgender women, as demonstrated in Chapter 1. For this reason, feminization of the face can be performed. The specific FFS procedures that change the shape of the forehead, nose, zygoma, chin and jaw region are respectively cranioplasty, rhinoplasty, zygoma osteotomy, genioplasty, and a mandibular angle reduction.[12,13] In this study only the forehead, nose, zygoma and chin region will be analysed. A separate shape analysis of the jaw and mandibular angle is not performed, due to the limited coverage of the facial template that is used to register the 3D surface scans to homologous 3D meshes.

The aim of this chapter is to analyse the morphology of different facial regions of transgender women prior to FFS. The morphology of the forehead, nose, chin and zygoma region is analysed by performing a separate PCA for each of the facial regions. Statistical analysis of the PC scores offers a comparison of the shape of the facial regions for transgender women, cisgender men and cisgender women. Analyzing separate facial regions allows for a more specific assessment of the shape differences between males and females. Furthermore, it offers the ability to assess specific facial regions of the faces of transgender women, which could be important within the context of FFS.

2.2. METHODS

The methods applied in this chapter are similar to the methods in Chapter 1. The same datasets are used, namely one dataset containing 1236 3D surface scans of cisgender male and female faces and one dataset containing 203 3D surface scans of the faces of transgender women. All scans in the dataset have been transformed into homologous 3D meshes following the preprocessing steps described in Chapter 1. The same aligned 3D meshes will be used in this chapter. However, this time the PCA and statistical analysis will not be performed on the entire face, but rather on specific facial regions.

Four different facial regions will be used, namely the forehead, nose, chin and zygoma. These regions are defined on the generic facial template, determining which vertices belong to which region. For each facial region, the correct vertices are extracted from the scans in the cisgender and transgender dataset. For the forehead, nose, chin and zygoma region this results in smaller meshes of respectively 1077, 603, 1482 and 1258 vertices. The four facial regions that will be analysed in this chapter are shown in figure 5.

For each facial region a separate PCA is performed on the x-, y- and z-coordinates of the 3D meshes of that specific region. Meaning only the coordinates of the extracted vertices will be used as input for the PCA. The scans from the cisgender dataset are used for the initial PCA, resulting in a set of PCs and PC scores for each of the four facial regions. Again, the PC scores for the scans in the transgender dataset are calculated using the covariance matrix from the initial PCA. Following the methods of Chapter 1, for each facial region the PCs that cumulatively account for 80% of the total variance within that region are used in further visualisations and statistical analysis. Statistical analysis is performed to identify the PCs that illustrate a difference between cisgender male, cisgender female and transgender female faces. All analyses will be carried out separately for the forehead, nose, chin and zygoma region.

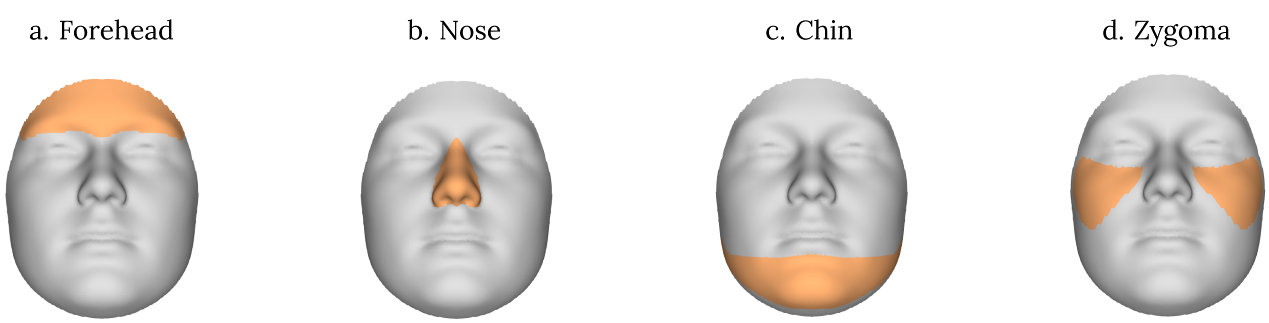

FIGURE 5. The regions used in this study visualized on the facial template: (a) Forehead, (b) Nose, (c) Chin and (d) Zygoma.

2.3. RESULTS

The PCs that cumulatively account for at least 80% of the shape variability are the first four PCs for the forehead and the nose region, and the first five PCs for the chin and zygoma region. An overview of the percentage of variability explained by the individual PCs of the four facial regions can be found in Appendix B.1. The results of the PCA and statistical analysis for the four facial regions will be separately discussed below.

2.3.1 Results for the forehead region

The shape variability captured by the first four PCs is visualized using the reconstruction of the forehead for which the PC scores differ three SDs from the average shape of the forehead, both in positive (+3SD) and negative (-3SD) direction. Distance maps are created showing the difference between the -3SD and +3SD reconstructed forehead regions in the direction perpendicular to the facial surface. A comprehensive overview of the visualizations of the first four PCs can be found in Appendix C.1. The distance maps for the PCs are shown in figure 6. From these visualisations the anatomical meaning of each PC is evaluated, as described in table 4.

TABLE 4. The anatomical meaning of the shape variability within the forehead region captured in the first four principal components.

PC	Main attributes in facial shape variability captured in each principal component
PC1	Projection of the brow ridges. A high PC score corresponds to a more pronounced forehead with prominent brow ridges, while a low PC score corresponds with a flatter, less pronounced forehead.
PC2	Shape of the superior part of the forehead, with higher PC scores resulting in a flatter forehead and lower PC scores in a more slanted forehead.
PC3	Shape of the central part of the forehead in relation to the lateral sides, with higher PC scores corresponding to a flatter forehead with an increased width and height.
PC4	Shape of the central part of the forehead in relation to the superior and inferior part of the brow ridge. Higher PC scores correspond to a flatter and less broad forehead.

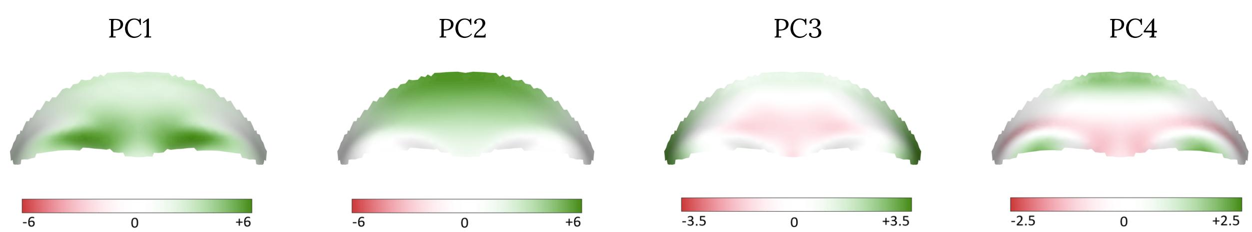

FIGURE 6. Distance maps illustrating for each of the principal components (PCs) the differences between the reconstructed forehead regions that differ three standard deviations from the average in the positive and in the negative direction. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

The descriptive statistics and distribution of the PC scores for the forehead region of cisgender men, cisgender women and transgender women can be found in Appendix C.2. A one-way MANOVA (CI=95%) of the first four PCs showed a significant difference in PC scores among the three groups, $F(8, 2868) = 109.884$, $p<.001$; Pillai's Trace = 0.469, partial $\eta 2 = 0.235$. Follow-up ANOVAs show a statistically significant difference among the three groups for all four PCs. However, PC2 has a low effect size (partial $\eta 2$) of .006 and will therefore be excluded from further analyses. The results are shown in table 5.

TABLE 5. Results of the ANOVA for the first four principal components.

PC	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η2
PC1	1679800.685	2	839900.343	192.358	.000	.211
PC2	14825.020	2	7412.510	4.562	.011	.006
PC3	297321.936	2	148660.968	191.703	.000	.211
PC4	20911.841	2	10455.920	16.959	.000	.023

Post hoc analysis is performed for PC 1, 3 and 4 using the Scheffe method, to offer individual comparisons between the groups of cisgender male, cisgender female and transgender female faces. The results are shown in table 6. A statistically significant difference ($p<.05$) between the groups of cisgender males and cisgender females is found in all three PCs. Transgender females compared to cisgender males show a statistically significant difference for PC3 ($p<.001$), and no significant difference for PC1 ($p=.852$) and PC4 ($p=7.33$). Finally, when comparing transgender females to the cisgender females, statistically significant differences ($p<.05$) are observed in all three PCs.

TABLE 6. Results of the post hoc analysis (Scheffe's method) to compare the PC scores of the forehead region for cisgender males, cisgender females and transgender females.

					95% Confidence Interval	
PC	Comparison between groups	Mean Difference	Std. Error	Sig.	Lower Bound	Upper Bound
PC1	Cis. Male - Cis. Female	67.434*	3.809	.000	58.100	76.767
	Trans. Female - Cis. Male	3.093	5.472	.852	-10.314	16.500
	Trans. Female - Cis. Female	70.527*	5.253	.000	57.656	83.397
PC3	Cis. Male - Cis. Female	-31.224*	1.605	.000	-35.158	-27.291
	Trans. Female - Cis. Male	13.381*	2.306	.000	7.731	19.031
	Trans. Female - Cis. Female	-17.843*	2.214	.000	-23.267	-12.419
PC4	Cis. Male - Cis. Female	-8.010*	1.431	.000	-11.518	-4.503
	Trans. Female - Cis. Male	1.620	2.056	.733	-3.418	6.658
	Trans. Female - Cis. Female	-6.390*	1.974	.005	-11.227	-1.554

* The mean difference is significant at the .05 level.

2.3.2 Results for the nose region

The shape variability captured in the first four PCs is visualized using the reconstruction of the nose for which the PC scores differ three SDs from the average nose, both in positive (+3SD) and negative (-3SD) direction. Distance maps are created showing the difference between the -3SD and +3SD reconstructed noses in the direction perpendicular to the facial surface. A comprehensive overview of the visualizations of the PCs can be found in Appendix D.1. The distance maps for the first four PCs are shown in figure 7. From these visualisations the anatomical meaning of each PC is evaluated, as described in table 7.

TABLE 7. The anatomical meaning of the shape variability within the nose region captured in the first four principal components.

PC	Main attributes in facial shape variability captured in each principal component
PC1	Forward projection of the nose, with higher PC scores corresponding to a larger, more protruding nose.
PC2	Length of the nose, with the length of the nose decreasing as the PC scores increase.
PC3	Asymmetry of the nose, with higher PC scores corresponding to a deviation of the nose to the left side, and lower PC scores corresponding to a deviation to the right side.
PC4	Shape of the profile of the nose. Higher PC scores correspond to a more convex nasal bridge, while lower PC scores correspond to a more concave nasal bridge. This also influences the position of the tip of the nose, with the tip being more lifted with lower PC scores.

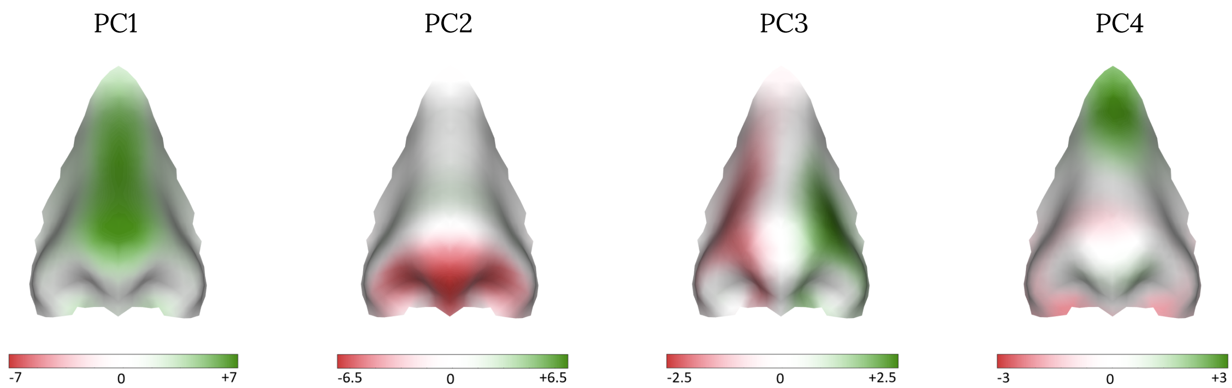

FIGURE 7. Distance maps illustrating for each of the principal components (PCs) the differences between the reconstructed nose regions that differ three standard deviations from the average in the positive and in the negative direction. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

The descriptive statistics and distribution of the PC scores for the nose region of cisgender men, cisgender women and transgender women can be found in Appendix D.1. A one-way MANOVA (CI=95%) on the first four PCs shows a significant difference in PC scores among the three groups, $F(8, 2868) = 42.213$, $p<.001$; Pillai's Trace = 0.211, partial $\eta 2 = 0.105$. Follow-up ANOVAs show a statistically significant difference among the three groups for PC 1, 2 and 4, and no statistically significant difference for PC3 ($p=.066$). PC2 is excluded from further analyses due to a low effect size (partial $\eta 2$) of .009. The results are shown in table 8.

TABLE 8. Results of the ANOVA for the first four principal components.

PC	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial $\eta 2$
PC1	409575.131	2	204787.566	99.614	.000	.122
PC2	22326.202	2	11163.101	6.575	.001	.009
PC3	1628.525	2	814.262	2.719	.066	.004
PC4	26942.913	2	13471.457	55.695	.000	.072

Post hoc analysis is performed for PC 1 and 4 to offer individual comparisons between the groups of cisgender male, cisgender female and transgender female noses. The results are shown in table 9. For both PCs a statistically significant difference (p<.001) is found between the cisgender males and cisgender females, as well as between the transgender females and the cisgender females. There is no statistically significant difference found between the transgender females and cisgender males for PC1 (p=.844) and PC4 (p=.084).

TABLE 9. Results of the post hoc analysis (Scheffe's method) to compare the PC scores of the nose region for cisgender males, cisgender females and transgender females.

					95% Confidence Interval	
PC	**Comparison between groups**	**Mean Difference**	**Std. Error**	**Sig.**	**Lower Bound**	**Upper Bound**
PC1	Cis. Male - Cis. Female	34.329*	2.6138	.000	27.924	40.733
	Trans. Female - Cis. Male	-2.188	3.754	.844	-11.387	7.012
	Trans. Female - Cis. Female	32.141*	3.604	.000	23.310	40.973
PC4	Cis. Male - Cis. Female	9.267*	.897	.000	7.070	11.464
	Trans. Female - Cis. Male	-2.867	1.288	.084	-6.023	.288
	Trans. Female - Cis. Female	6.399*	1.2366	.000	3.370	9.429

* The mean difference is significant at the .05 level.

2.3.3 Results for the chin region

The shape variability captured in the first five PCs is visualized using the -3SD and +3SD reconstructions and the distance maps showing the difference between the -3SD and +3SD reconstructed noses in the direction perpendicular to the facial surface. A comprehensive overview of the visualizations of the PCs can be found in Appendix E.1. The distance maps for the first five PCs are shown in figure 8. From these visualisations the anatomical meaning of each PC is evaluated, as described in table 10.

TABLE 10. The anatomical meaning of the shape variability within the chin region captured in the first five principal components.

PC	Main attributes in facial shape variability captured in each principal component
PC1	Length of the chin, with high PC scores representing a shorter chin.
PC2	Forward projection of the chin, with higher PC scores resulting in a more protruding chin.
PC3	Outline of the jaw, with higher PC scores resulting in a slimmer jaw and and high PC scores resulting in a broader, more squared jaw.
PC4	Shape of the chin, with high PC scores representing a rounder chin and low PC scores representing a more squared chin.
PC5	Changes in the lateral sides of the region, corresponding to the overall width of the jaw. High PC scores correspond with a slimmer jaw.

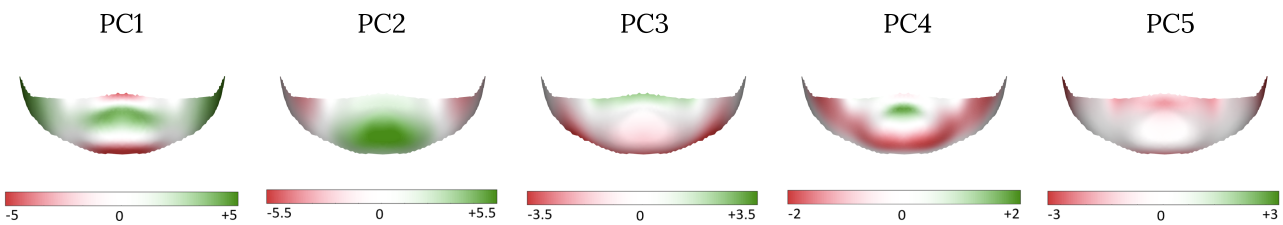

FIGURE 8. Distance maps illustrating for each of the principal components (PCs) the differences between the reconstructed chin regions that differ three standard deviations from the average in the positive and in the negative direction. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

A one-way MANOVA (CI=95%) showed a significant difference in PC scores among cisgender men, cisgender women and transgender women, $F(10, 2866) = 49.042$, $p<.001$; Pillai's Trace = 0.292, partial $\eta 2$ = 0.146. Follow-up ANOVAs show a statistically significant difference among the three groups for PC 2, 3, 4 and 5, with no statistically significant difference for PC1 ($p=.210$). PC3 is excluded from further analyses due to a low effect size (partial $\eta 2$) of .009. The results are shown in table 11. The descriptive statistics and distribution of the PC scores for the different groups can be found in Appendix E.2.

TABLE 11. Results of the ANOVA for the first four principal components.

PC	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial $\eta 2$
PC1	19463.281	2	9731.641	1.564	.210	.002
PC2	522823.702	2	261411.851	104.325	.000	.127
PC3	11038.862	2	5519.431	6.266	.002	.009
PC4	82867.263	2	41433.631	52.478	.000	.068
PC5	61228.432	2	30614.216	64.952	.000	.083

Results of the post hoc analysis of PC 2, 4 and 5 are shown in table 12. A statistically significant difference between cisgender male and female faces is found for PC 2 and 5, with no statistically significant difference in PC4 ($p=.495$). For the transgender female and cisgender male group, there is a statistically significant difference found in PC 2 and 4, with no statistically significant difference in PC5 ($p=.970$). When comparing transgender females to cisgender females the is a statistically significant difference found for all three PCs.

TABLE 12. Results of the post hoc analysis (Scheffe's method) to compare the PC scores of the chin region for cisgender males, cisgender females and transgender females.

PC	Comparison between groups	Mean Difference	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
PC2	Cis. Male - Cis. Female	40.929*	2.886	.000	33.858	48.000
	Trans. Female - Cis. Male	-13.408*	4.145	.005	-23.565	-3.252
	Trans. Female - Cis. Female	27.521*	3.979	.000	17.771	37.271
PC4	Cis. Male - Cis. Female	-1.923	1.620	.495	-5.892	2.046
	Trans. Female - Cis. Male	22.770*	2.327	.000	17.069	28.472
	Trans. Female - Cis. Female	20.848*	2.234	.000	15.375	26.321
PC5	Cis. Male - Cis. Female	-12.919*	1.252	.000	-15.986	-9.852
	Trans. Female - Cis. Male	-.441	1.798	.970	-4.846	3.964
	Trans. Female - Cis. Female	-13.360*	1.726	.000	-17.588	-9.131

* The mean difference is significant at the .05 level.

2.3.4 Results for the zygoma region

Finally, the shape variability captured in the first five PCs is visualized using distance maps illustrating the difference between the -3SD and +3SD reconstructed zygoma regions in the direction perpendicular to the facial surface. A comprehensive overview of the visualizations of the first five PCs can be found in Appendix F.1. The distance maps for the first five PCs are shown in figure 9. From these visualisations the anatomical meaning of each PC is evaluated, as described in table 13.

TABLE 13. The anatomical meaning of the shape variability within the zygoma region captured in the first five principal components.

PC	Main attributes in facial shape variability captured in each principal component
PC1	Volume of the cheeks, with higher PC scores representing more volume.
PC2	Shape of the medial part of the region in relation to the lateral part of the region. For higher PC scores, the medial part is tilted outwards and the lateral parts inward.
PC3	Projection and change in shape in the superior part of the region, with higher PC scores representing more projection.
PC4	Shape of the lateral corners of the regions in relation to the medial inferior part of the region.
PC5	Lateral outline of the zygoma region. High PC scores are representing an extension in lateral direction.

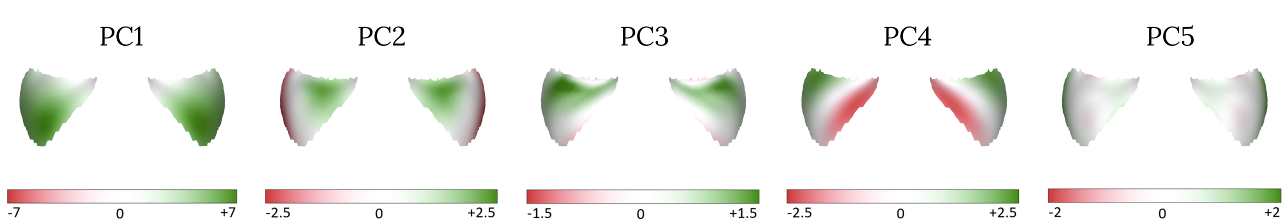

FIGURE 9. Distance maps illustrating for each of the principal components (PCs) the differences between the reconstructed zygoma regions that differ three standard deviations from the average in the positive and in the negative direction. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

The one-way MANOVA (CI=95%) for the first five PCs showed a significant difference in PC scores among the three groups (cisgender male, cisgender female and transgender female), $F(10, 2866) = 94.973$, $p<.001$; Pillai's Trace = 0.498, partial $\eta 2$ = 0.249. Follow-up ANOVAs show a statistically significant difference among the three groups for all five PCs. However, PC4 is excluded from further analyses due to a low effect size (partial $\eta 2$) of .010. The results are shown in table 14. The descriptive statistics and distribution of the PC scores for the different groups can be found in Appendix F.2.

TABLE 14. Results of the ANOVA for the first five principal components.

PC	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial $\eta 2$
PC1	1629221.240	2	814610.620	314.213	.000	.304
PC2	189379.544	2	94689.772	50.131	.000	.065
PC3	111808.059	2	55904.029	72.624	.000	.092
PC4	5693.999	2	2846.999	7.010	.001	.010
PC5	14688.240	2	7344.120	25.164	.000	.034

A post hoc analysis is performed for PC 1, 2, 3 and 5. The results are shown in table 15. When comparing the zygoma region for cisgender males and females, a statistically significant difference is found for PC 1, 2 and 3, with PC5 showing no statistically significant difference (p=.364). For the transgender female and cisgender male group, there is a statistically significant difference found in PC 1, 3 and 5, with no statistically significant difference in PC2 (p=.983). When comparing transgender females to cisgender females there is a statistically significant difference found for all four PCs.

TABLE 15. Results of the post hoc analysis (Scheffe's method) to compare the PC scores of the zygoma region for cisgender males, cisgender females and transgender females.

					95% Confidence Interval	
PC	Comparison between groups	Mean Difference	Std. Error	Sig.	Lower Bound	Upper Bound
PC1	Cis. Male - Cis. Female	-73.127*	2.935	.000	-80.320	-65.935
	Trans. Female - Cis. Male	31.750*	4.216	.000	21.419	42.081
	Trans. Female - Cis. Female	-41.377*	4.047	.000	-51.295	-31.460
PC2	Cis. Male - Cis. Female	-22.753*	2.505	.000	-28.892	-16.614
	Trans. Female - Cis. Male	-.665	3.599	.983	-9.484	8.153
	Trans. Female - Cis. Female	-23.418*	3.455	.000	-31.883	-14.953
PC3	Cis. Male - Cis. Female	11.295*	1.599	.000	7.376	15.214
	Trans. Female - Cis. Male	13.959*	2.297	.000	8.330	19.588
	Trans. Female - Cis. Female	25.259*	2.205	.000	19.850	30.658
PC5	Cis. Male - Cis. Female	1.401	.985	.364	-1.012	3.814
	Trans. Female - Cis. Male	8.178*	1.415	.000	4.712	11.644
	Trans. Female - Cis. Female	9.579*	1.358	.000	6.251	12.906

* The mean difference is significant at the .05 level.

2.4. DISCUSSION

The aim of this chapter was to analyse the facial morphology of transgender women for four separate facial regions; the forehead, nose, chin and zygoma. The morphology of each of these facial regions of transgender women is analysed and compared to cisgender men and women. For this, four separate PCAs and statistical analyses are performed.

2.4.1 Interpretation of the results

Starting with the forehead area, all four PCs showed a statistically significant difference among the groups of cisgender men, cisgender women and transgender women. However, PC2 showed a low effect size, suggesting that the difference in the shape of the superior part of the forehead will likely have limited practical or clinical significance. For PC 1, 3 and 4 a statistically significant difference was found between cisgender men and cisgender women, which correlates with existing literature describing the brow ridge of females to be less protruding with the medial part of the forehead placed more anteriorly.[46] The shape of the forehead of transgender women in terms of the projection of the brow ridges and central part of the forehead differs from the shape of the forehead of cisgender women, and can for the most part be considered as masculine. This supports the wish for feminization of the forehead for transgender women.

For the nose region, only PC 1 and 4 showed both a statistically significant difference and a sufficient effect size. This means that there is no significant difference observed in the length (PC2) and asymmetry of the nose (PC3). For PC 1 and 4 a statistically significant difference is observed between cisgender men and women, as well as between transgender women and cisgender women. No significant difference is observed between transgender women and cisgender men. This suggests that the shape of the nose of transgender women corresponds to the shape of the nose of cisgender men in terms of forward projection (PC1) and shape of the profile of the nose (PC4). This study illustrates that cisgender men and transgender women show a larger forward projection of the nose in comparison to cisgender women. Furthermore, the nose of cisgender men and transgender women tend to have a more convex nasal bridge, while cisgender women tend to have a more concave nasal bridge with a more lifted tip. This is in accordance with the features of the nose that have been described as being sexually dimorphic.[46,30] Apart from the projection and shape of the nose, the overall size and width of the nose has been described as a sexually dimorphic feature. From the visualizations and the attributed anatomical meanings of the PCs it can be observed that the overall width of the nose is not represented within one of the first four PCs. This means that within the facial shape variation of the nose, the width did not explain enough of the variability to be represented within the PCs that cumulatively account for 80% of the shape variability. Therefore, possible differences in width of the nose are not analysed within this research.

For the analysis of the shape of the chin, the first five PCs are used. No statistically significant difference was found for PC1, suggesting there is no difference in the length of the chin between cisgender men, cisgender women and transgender women. For PC3 the statistical analysis did show a statistically significant difference, but due to the low effect size this difference is likely to have limited practical significance. This PC represents the shape variability in the outline and shape of the jaw. Based on the anatomical meaning of the PCs a difference would have been expected, as literature describes both the length of the chin and the shape of the jaw as features illustrating sexual dimorphism.[15,46] Furthermore, men tend to have more protruding chins, a more squared shape of the chin and a broader jaw.[15,47] However, when comparing cisgender males to cisgender females, this study only shows a difference in forward projection of the chin (PC2) and width of the jaw (PC5). The shape of the chin

(PC4) did not show a statistically significant difference between cisgender men and women. When comparing the PC scores of transgender women to the PC scores of cisgender men and women, a statistically significant difference is found for PC 2, 4, and 5. This suggests that the shape of the chin and jaw of transgender women does significantly differ from cisgender women, explaining the need for feminization of the chin and jaw. However, the PC scores of transgender women also show a statistically significant difference in comparison to the PC scores of cisgender men for PC 2 and 4, suggesting that the projection and shape of the chin of transgender women differ from both cisgender males and females.

Lastly, the results for the zygoma region are interpreted for the first five PCs. Statistically significant difference was found among the groups of cisgender males, cisgender females and transgender females for all five PCs. However, the observed difference for PC4 holds little practical significance, as indicated by the low effect size. This PC represents the shape of the lateral corners of the regions in relation to the medial inferior part of the region. A statistically significant difference between cisgender male and female faces is found for PCs 1, 2 and 3, suggesting a difference in volume of the cheeks and the projection in the medial and superior part of the region. This is largely in accordance with the described features of sexual dimorphism within this facial region. Female tend to show more volume of the cheeks with a more pronounced zygomatic prominence.[15,48] Statistically significant differences were also found between transgender women and cisgender women for PC 1, 2, 3 and 5, and between transgender women and cisgender men for PC 1, 3 and 5. This suggests that the shape of the zygomatic region of transgender women does not only differ from the typically female shape, but also from the typically masculine shape. One of the facial features that is described to change due to the effects of gender-affirmative hormone treatment for transgender women is the volume of the cheeks.[31] This could explain the difference between transgender women and cisgender men, especially for PC1. Another attributing factor could be the possible effects of dermal fillers within the zygoma region.[49]

2.4.2 Study limitations

As the study presented in this chapter largely follows the methods of Chapter 1 a lot of the limitations described in the previous chapter apply to this study as well. This includes the choice of threshold for the cumulative explained variability within the dataset. In both chapters a threshold of 80% explained variability is used. For the full facial shape this required 12 PCs, while for the different facial regions only 4 or 5 PCs were required. In comparison to the entire facial shape, the shapes of the separate facial regions are less complicated and the shape variability can be captured with less PCs. However, it should be considered to use a higher threshold for the PCA of the separate facial regions, as the additional PCs still explain a considerable percentage of the explained variability. As mentioned above, for the nose region it could be observed that by limiting the amount of PCs to explain 80% of the variability, an important part of the shape variability was not included in the analysis.

Other limitations that apply to both studies are the limited generalizability to different races, the selection bias within the transgender dataset, and the possible inaccuracies in the capturing and preprocessing of the 3D surface scans. The problem with insufficient coverage of the facial template for the full facial shape only influences the forehead and chin region. The regions of the nose and zygoma are not influenced by this limitation, as these regions do not extend to the edges of the template.

Furthermore, a different definition of some of the facial regions should be considered. The region that is defined as the chin region, does not only cover the surface area of the chin, but continues to the lateral sides of the face. The reason for this was the ability to include some additional information about the width and size of the jaw. Due to the limited coverage of the facial template for the jaw and mandibular angle, the jaw could not be analysed as a separate region. However, it might result in a more reliable analysis to limit the region to the area of the chin. In future research a facial template with a better coverage of the area around the jaw should be used, allowing for separate analysis of the chin and the jaw region. For the zygoma region it should be considered whether the region should include the full area of the cheeks, or should be limited to a smaller region more specific for the zygomatic prominence. With the definition for the zygoma area as used in this study, the region does not completely extend to the lateral edges of the facial template. This might exclude some of shape variation related to the lateral projection of the zygoma.

Finally, there are still some limitations with the proposed method to analyse the facial shape by comparing each PC individually. As mentioned before, the overall facial shape, or in this case the overall shape of the facial regions, is defined by the combination of all PCs. It is therefore not possible to attribute the change of one single facial feature or facial region to one single PC. By analyzing the facial regions separately, the shape analysis is simplified as the facial shape variation can be described by a smaller set of PCs that are specific for the facial region. This limits the problem of PCs describing shape variability in different regions of the face. However, it should still be considered that analyzing the facial shape differences in terms of the separate PCs might offer limited information about the overall shape of the facial regions. This also makes it difficult to give a overall conclusion about the masculinity or femininity of the face or facial regions. It would therefore be of added value to be able to combine the PC scores of the different PCs into a single score that captures the overall facial shape in comparison to male and female facial shape. The possibility of creating such a masculinity/femininity score is explored in chapter 4.

2.5. CONCLUSION

The research presented in this chapter focused on the geometrical shape analysis of the forehead, nose, chin and zygoma region. Separate PCAs and statistical analyses are performed to identify the PCs that illustrate a difference in PC scores between transgender women, cisgender men and cisgender women. Analyzing separate facial regions allows for a more specific assessment of the shape differences, as the problem of individual PCs describing facial shape variation across different regions of the face is limited. Part of the PCs of each of the facial regions showed statistically significant differences in shape between cisgender men and women, as well as between transgender women and cisgender women. Large part of these differences correspond to existing literature describing sexual dimorphism. The differences in shape between transgender women and cisgender women illustrate the need for facial feminization surgery. Same as in Chapter 1, for the separate facial regions there are also some differences observed between transgender women and cisgender men. These differences might be an indication of the effects of gender-affirmative hormone treatment. However, further research into the effects of hormone treatment is needed to be able to collaborate on the exact nature of these differences. Finally, it should be mentioned that although the method proposed in this chapter can offer more specific information about the morphology of different facial regions, it is difficult to give an overall conclusion about the masculinity or femininity of the shape of specific facial regions. Therefore, further research should be focused on the possibility of combining the PC scores of the different PCs into one single masculinity/femininity score.

Chapter 3. The effect of facial feminization surgery on the facial shape; Comparing preoperative and postoperative 3D surface scans of transgender women

3.1. INTRODUCTION

The research presented so far has been focused on the facial morphology of transgender women prior to FFS. In the previous two chapters, the shape of the face and facial regions of transgender women is described using a PCA-approach and compared to the facial shape of cisgender men and women. This offers more insight into the indications of FFS. However, the aim of the overall study is to look at both the facial morphology prior to FFS, as well as after FFS. Therefore, within this chapter the changes in facial morphology that are achieved in FFS are evaluated.

An evaluation of the facial morphology before and after FFS can illustrate the exact changes in facial shape that are obtained during FFS. As explained before, FFS comprises of different craniomaxillofacial procedures with the aim to change masculine facial features into more feminine facial features. Examples of procedures that can be part of FFS are cranioplasty, rhinoplasty, genioplasty, zygoma osteotomy and mandibular angle reduction.[12,13] Although during these procedures most modifications are made to the shape of the skull of the patient, the overall result of FFS is based on the shape of the face. Therefore, preoperative and postoperative facial 3D surface scans can be used to analyse the effects of FFS.

Evaluation of the facial shape before and after FFS can be used to quantify the effect of the procedure. By comparing both the preoperative and postoperative facial shape of transgender women to the facial shape of cisgender women, it can be evaluated whether FFS has resulted in a more feminine facial shape. This allows for a quantification of the results of FFS, which can be of added value in monitoring the quality of FFS procedures, as well as allow for a comparison between different surgical techniques. Furthermore, more insight into the exact changes in facial morphology achieved in different FFS procedures can help patients form realistic expectations for the FFS procedures.

The study presented in this chapter focuses on comparing the facial morphology of transgender women prior to FFS with the facial morphology after FFS. This is done by comparing the preoperative and postoperative facial 3D surface scans in terms of the average change in coordinates of the vertices, as well as using a PCA-approach similar to the methods described in Chapter 1 and 2. To be able to evaluate the effects of specific FFS procedures, the analyses are performed for separate facial regions. The evaluation of the change in facial shape can offer more information about the exact effect of different FFS procedures. This can be of added value for evaluating the results of FFS, as well as give patients a better idea of what to expect after FFS procedures.

3.2. METHODS

3.2.1 Dataset

The dataset used in this study consists of facial 3D surface scans of transgender women, who have undergone FFS at the Amsterdam UMC (Amsterdam, The Netherlands). The scans are obtained using the Artec Leo handheld 3D scanner (Artec 3D, Luxembourg, Luxembourg) and converted to 3D meshes using the Artec Studio Professional software (Artec 3D, Luxembourg, Luxembourg). In total 28 Caucasian transgender women aged 18-50 years without known genetic disorders or craniofacial dysmorphism are included in the dataset. For each individual, both a preoperative and postoperative facial 3D scan is available.

3.2.2 Preprocessing

The preprocessing of the facial 3D surface scans follows the same methods as in Chapter 1 and 2. Each scan in the dataset is represented as a homologous 3D mesh and aligned using a Generalized Procrustes Analysis. The four facial regions (forehead, nose, chin and zygoma) are extracted from the 3D meshes following the methods described in Chapter 2. This results in 28 preoperative and 28 postoperative 3D meshes for each of the facial regions. However, as FFS comprises a range of different craniomaxillofacial procedures, not all individuals in the dataset have undergone the same procedure or combination of procedures. For each facial region, only the 3D meshes of individuals that have undergone an FFS procedure that addresses that specific facial region will be included. The remaining 3D meshes of the facial regions are removed from the dataset. Table 16 gives an overview of the number of individuals within the dataset that undergone the different FFS procedures affecting the forehead, nose, chin and zygoma region.

TABLE 16. Overview of the number of individuals within the dataset that undergone the specified FFS procedures.

FFS procedure	Affected facial region	Number of individuals
Cranioplasty	Forehead	22
Rhinoplasty	Nose	25
Genioplasty	Chin	13
Zygoma osteotomy	Zygoma	8

3.2.3 Principal component scores and Z-scores

The shapes of the facial regions are described using PC scores. Chapter 2 described the PCA of the cisgender dataset for the different facial regions. For each facial region a separate PCA was performed on the x-, y- and z-coordinates of the 3D meshes. The resulting covariance matrices of these separate PCAs are now used to calculate the PC scores for the 3D meshes in the transgender dataset. For each facial region, the scores are calculated using the PCA specifically performed for that facial region. Again, the PC scores of the PCs that cumulatively account for 80% of the shape variability within the cisgender dataset are used. For the forehead and nose region this results in the first four PCs and for the chin and zygoma region this results in the first five PCs.

The PC scores of the preoperative and postoperative scans are translated to Z-scores. The Z-score indicates how many standard deviations the PC score deviates from the mean of the reference distribution. In this case the PC scores of transgender women are compared to the distribution of PC scores of the cisgender women. The Z-scores thus represent how many standard deviations the PC score deviates from the cisgender female mean shape. The average preoperative and postoperative Z-scores are calculated for the different PCs and facial regions. The preoperative to postoperative change in Z-score relative to the mean is calculated to evaluate whether the shape of the facial region has moved closer to the mean female shape, due to the changes made during FFS.

3.2.4 Visualization of the average change in facial shape

A different way to analyse the change in facial shape resulting from FFS is to look at the position of the x-, y- and z-coordinates of the 3D meshes. The shape of the facial regions is described using homologous 3D meshes for which the position of the vertices correspond between the different 3D meshes. Meaning, corresponding vertices are expected to be located in the same position on the facial surface. The change in facial shape can thus be described by the change in coordinates of the vertices. The average change in the x-, y- and z-coordinate from the preoperative to the postoperative 3D meshes is calculated for each vertex in the 3D mesh and translated to the average change in the direction of the surface normal. These calculations are performed for each of the four facial regions. Distance maps are created to illustrate the differences between the preoperative and postoperative scans. The average change in the direction perpendicular to the facial surface is visualized using a red-to-green color scale on the mean shapes of the forehead, nose, chin and zygoma region. These distance maps illustrate the average change resulting from respectively cranioplasty, rhinoplasty, genioplasty and zygoma osteotomy.

3.3. RESULTS

The PC scores are calculated for the preoperative and postoperative 3D surface scans and represented as Z-scores. This is done for the different PCs of the forehead, nose, chin and zygoma region. The change in Z-score relative to the mean illustrates whether the average PC score has moved closer or further away from the mean female PC score for that specific PC and facial region. Table 17 shows the average Z-scores for the preoperative and postoperative 3D meshes, as well as the change in Z-score relative to the mean female shape for each PC.

The distance maps illustrated in figure 10 show the average change from the preoperative to postoperative 3D meshes for the corresponding vertices. The average change is visualized in the direction perpendicular to the facial surface using a red-to-green color scale.

3.4. DISCUSSION

The study presented in this chapter focused on analysing the difference in facial morphology between preoperative and postoperative 3D surface scans of transgender women. The change in shape of four different facial regions is described as Z-scores and as the average change in coordinates of the corresponding vertices.

TABLE 17. Overview of the calculated Z-scores for the preoperative and postoperative 3D meshes, for the different facial regions and PCs.

Facial region	N	PC	Average Z-score preoperative	Average Z-score postoperative	Change in Z score relative to the mean
Forehead	22	PC1	1.593	-0.668	-0.926
		PC2	0.140	0.770	+0.630
		PC3	-0.204	0.783	+0.579
		PC4	-0.116	-0.089	-0.027
Nose	25	PC1	0.976	-0.550	-0.426
		PC2	-0.002	-0.543	+0.541
		PC3	0.065	-0.071	+0.006
		PC4	0.778	0.084	-0.695
Chin	13	PC1	-0.424	-0.605	+0.181
		PC2	0.920	0.487	-0.433
		PC3	-0.250	0.541	+0.290
		PC4	0.654	-0.498	-0.156
		PC5	-1.089	-0.315	-0.775
Zygoma	8	PC1	-2.105	-0.627	-1.478
		PC2	0.129	0.316	+0.187
		PC3	0.865	-0.108	-0.757
		PC4	-0.039	0.347	+0.307
		PC5	0.549	-0.483	-0.067

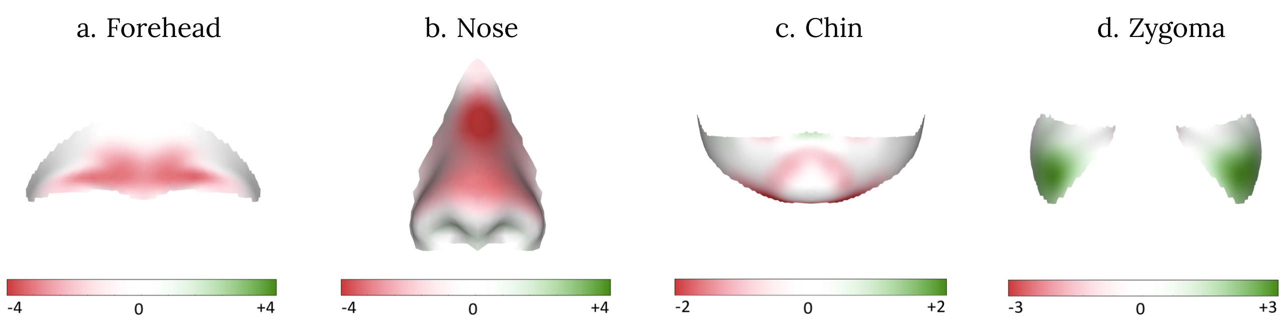

FIGURE 10. Distance maps illustrating the average distance between corresponding vertices of the preoperative and postoperative 3D meshes of the different facial regions. The distances (in mm) are visualised in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

3.4.1 Interpretation of the results

When looking at the average Z-scores it can be observed that there is a lot of variance between the average Z-scores for both the preoperative and postoperative 3D meshes. For the preoperative scans, some PCs illustrate an average deviation from the female mean of more than 1 SD. This is the case for PC1 of the forehead region, PC5 of the chin region and PC1 of the zygoma region. For each of these PCs, the change in Z-score shows that the postoperative average PC score has moved closer to the female mean. For the postoperative scans, the highest absolute average Z-score is observed for PC3 of the forehead region, which has a Z-score of 0.783. This means that the average Z-score for all PCs now lie within 1 SD from the mean cisgender female shape. However, as can be observed in table 17 a lot of the PC scores have also move further away from the female mean when comparing the preoperative and postoperative Z-scores. The PC scores were translated to Z-scores as a way to correlate the facial shape of the transgender women to the facial shape of the cisgender women. However, this method has some important limitations, which make it difficult to interpret the results within the context of FFS. These limitations will be further discussed in the next section.

The distance map of the forehead region shows a reduction of the volume of the brow ridge as well as in the medial region of the forehead. This corresponds with the expected change in facial shape after a cranioplasty.[48] The distance map of the nose region also displays the expected change achieved during a feminising rhinoplasty. The general size of the nose is reduced, with a large change to the nasal bridge to create a more concave shape. For the chin region the change in shape is less evident, but does show a refinement in terms of the width of the chin. The shape of the chin becomes more rounded as opposed to squared, which correlates to the changes induced during genioplasty.[48] Within the zygoma region an increase in volume can be observed in the distance map. When compared to the expected results of the zygoma osteotomy, the increase in volume would have been expected to be located more laterally.

3.4.2 Study limitations

Within this chapter, the Z-score is used as a measure for how feminine the face is. However, the Z-score only illustrates how many SDs the PC score deviates from the cisgender female mean. It does not take into account whether a face deviates from the mean because it is not feminine enough, or because it is more feminine than the average female face. The aim of FFS is not always to create facial features that correspond to the 'average' female face, but rather to a face that can be considered as 'extra feminine' to eliminate any external confusion about the gender of the patient in terms of facial features. A deviation from the mean is therefore not necessarily a bad thing. Furthermore, in Chapter 1 and 2 it was already demonstrated that by analyzing the PC scores for each PC separately it is difficult to give an overall conclusion about the masculinity or femininity of the entire shape of the face or facial region. A possible solution to both these problems would be to combine the PC scores of the different PCs into one score that represents the masculinity/femininity of the face or specific facial region. The possibility to create and evaluate such a masculinity/femininity score will be explored in Chapter 4.

Another important limitation of this study is that the assumption is made that the change in facial shape between the preoperative and postoperative 3D surface scan is solely dedicated to FFS. However, several other factors can influence the facial shape, especially with a longer time period between the two scans. Changes in facial shape over time can for example be attributed to a change in BMI, the possible effects of gender-affirmative hormone therapy or the presence of postoperative swelling. The influence of these factors on the facial shape within this study is difficult to determine, due to the limited number of included scans and the absence of additional data such as weight at the time of the scan. However, it would be possible to explore the effects of these factors on the facial shape in future research using 3D shape analysis. The change in facial shape due to the factors described above can be partially minimised, by reducing the time between the preoperative and postoperative 3D surface scan. Part of the preoperative scans have been made quite a long time before the FFS was performed, resulting in relatively longer time period between the preoperative and postoperative 3D surface scan. The time between the FFS procedure and the postoperative scan should not be reduced, due to the presence of postoperative swelling. However, it would allow for more reliable comparisons if all postoperative scans would be made after the same time period following surgery.

The observed change in facial shape can also be attributed to a difference in alignment between the preoperative and postoperative 3D surface scans. Both the preoperative and the postoperative

3D surface scans are aligned to the facial template using the Generalized Procrustus Algorithm (GPA). Changes in facial shape as a result of FFS can result in a slightly different alignment for the two scans. A slight translation or rotation between the two 3D meshes can increase or decrease the observed difference in shape in the facial regions. This effect can be minimized by aligning the preoperative and postoperative 3D surface scans based on only the facial regions or structures that are not changed during FFS.

Lastly, a large limitation of the study presented in this chapter is the relatively small number of included scans. Especially for the evaluation of the genioplasty and zygoma osteotomy the limited number of scans might lead to unreliable results. Due the large individual variety and the small number of scans, the results of the study are difficult to be interpreted and translated to the transgender population. However, as more patients are being operated, the number of available preoperative and postoperative 3D surface scans keeps increasing and more data can be added to improve the reliability of the results of the study.

3.5. CONCLUSION

The aim of this chapter was to compare the facial morphology of transgender women prior to FFS with the facial morphology after FFS. The changes in shape for the forehead, nose, chin and zygoma region are analysed using two different approaches. In the first approach, the PC scores of the facial regions are calculated and described as Z-scores. These Z-scores illustrate how many SDs the PC scores of transgender women deviate from the mean of the PC scores of cisgender women. However, due to the relatively small number of scans and limitations of the proposed method, the results were inconclusive and of limited use to quantify the effect of the different FFS procedures. The second approach compares the preoperative and postoperative facial 3D surface scans in terms of the average change in coordinates of the corresponding vertices. The average change in position of the vertices is displayed as a distance map, illustrating the changes in morphology achieved by the different FFS procedures. The illustrated changes largely correlate with the changes expected for the cranioplasty, rhinoplasty and genioplasty. For the zygoma osteotomy an increase of volume within the region was observed, but was expected to be located more laterally. Although the proposed method to illustrate the average change in facial shape based on vertex coordinates might add some insight to the changes achieve in FFS, the proposed method for quantifying the facial shape using Z-scores was not useful within the context of FFS. Quantifying the effect of different FFS procedures on the facial morphology can be of added value for evaluating the results of FFS, as well as give patients a better idea of what to expect after FFS procedures. Combining the PC scores of different PCs into a single score that represents the masculinity/femininity of the face could offer a better approach to quantifying the facial shape. By basing this score on both the cisgender male and cisgender female distribution of PC scores, the limitations of the method used in this chapter can be overcome.

Chapter 4. Scoring the masculinity or femininity of the face of transgender women before and after facial feminization surgery based on principal component analysis

4.1. INTRODUCTION

One of the largest limitations of the studies presented in the chapters above is the limited use of analyzing the PCs separately. Although evaluating the PC scores of each PC separately did offer some insight into which facial features differ between cisgender male, cisgender female and transgender female faces, individual PCs describe only part of the facial shape. The full facial shape or shape of the separate facial regions can only be described by the combination of a sufficient number of PCs. To evaluate the degree of masculinity or femininity of the face or facial regions it is therefore needed to look at the combination of the PC scores instead of the individual PC scores.

The ability to determine the degree of masculinity or femininity of the face has been explored in previous studies. However, most of these studies are based on 2D images or geometric measurements of the 3D facial shape.[50,51,52] As 2D measurements offer limited information about the complex facial shape, the use of 3D shape analysis could provide a more comprehensive assessment of the facial morphology. As the full facial shape of an individual can be described by a relatively small amount of PCs, using the PCs could offer a valuable basis for scoring the degree of masculinity and femininity of the face.

Being able to score the degree of masculinity or femininity present within the facial shape can be of important value within the field of FFS. The indications for FFS are very subjective and difficult to quantify. Partially because the indications are based on the dysphoric feelings of the individual patient, but also because it is difficult to quantify which parts of the face show the most masculine features. This is where a masculinity/femininity score could be of added value. Another possible application of a masculinity/femininity score within the context of FFS would be to evaluate the results of the different FFS procedures. As the aim of FFS is to create a more feminine appearance of the face, it would be expected that when scoring the faces of transgender women preoperative and postoperative the masculinity/femininity score changes from a more masculine score to a more feminine score. Evaluating the masculinity/femininity scores of the postoperative facial shape could illustrate whether the desired effect has been achieved or whether follow-up procedures might be necessary. To quantify the indications for different facial regions, as well as accurately evaluate the results of different FFS procedures, calculating independent masculinity/femininity scores should be possible for the forehead, nose, chin and zygoma region.

The aim of the research presented in this chapter is to develop and evaluate a method to score the femininity or masculinity of an individual face or facial region, based on the PC scores of cisgender males and females. The presented masculinity/femininity score is calculated for the faces of transgender women before and after FFS. Being able to evaluate how masculine or feminine the face or specific facial regions of transgender women are prior to FFS, could help in determining which FFS procedures should be performed. Furthermore, comparing the masculinity/femininity scores before and after FFS can be used to illustrate whether a feminization of the face is achieved.

4.2. METHODS

4.2.1 Dataset

Two different datasets of PC scores are used within this study, a cisgender dataset and a transgender dataset. The masculinity/femininity score will be based on the distribution of the PC scores of the male and female cisgender faces. For this purpose, the cisgender dataset that has been described in Chapter 1 and 2 is used. This dataset contains the 3D surface scans of 1236 cisgender male and female faces. The PC scores resulting from the PCA of the entire face (Chapter 1) and the separate facial regions (Chapter 2) are used within this study as a base to calculate the masculinity/femininity score.

The dataset of preoperative and postoperative 3D surface scans of transgender women will be used to evaluate the use of the masculinity/femininity score in the context of FFS. This dataset has been described in Chapter 3. The PC scores describing the shape of the face and separate facial regions will be used. For the full facial region, the PC scores of the preoperative and postoperative 3D surface scans of all 28 individuals in the dataset will be used. These individuals have undergone different combinations of FFS procedures. For the separate facial regions, only the individuals who have undergone an FFS procedure regarding that specific facial region are included within the study. For the forehead, nose, chin and zygoma region, the PC scores of respectively 22, 25, 13 and 8 preoperative and postoperative 3D surface scans are used. This corresponds with the subsets of the transgender dataset as described in Chapter 3.

4.2.2 Calculating the masculinity/femininity score

The calculation of the masculinity/femininity score is based on how the PC scores for the different PCs compare to the PC scores of cisgender male and female faces. Unlike in the previous chapters, the PCs are not analyzed independently, but are combined into one single score. As each individual face can be described by a combination of n PC scores, this means that each face can be considered as a datapoint located somewhere in an n-dimensional space. The location of each face within the n-dimensional space is determined by the values of the PC scores. By calculating the masculinity/femininity score, the dimensionality of the data will be reduced from n dimensions to one single value representing the masculinity or femininity of the face.

First, for each included PC, the mean of the PC scores of the cisgender male faces and the mean of the PC scores of the cisgender female faces are calculated. The combinations of the mean PC scores for the first n PCs describe the mean male facial shape and the mean female facial shape. These two facial shapes can be considered as two datapoints located in the n-dimensional space as described before. For each of the PCs a vector can be drawn from the mean female PC score to the mean male PC score. For all PCs combined, this results in an n-dimensional vector that runs from the mean male facial shape to the mean female facial shape. This vector represents the direction in the n-dimensional space explaining

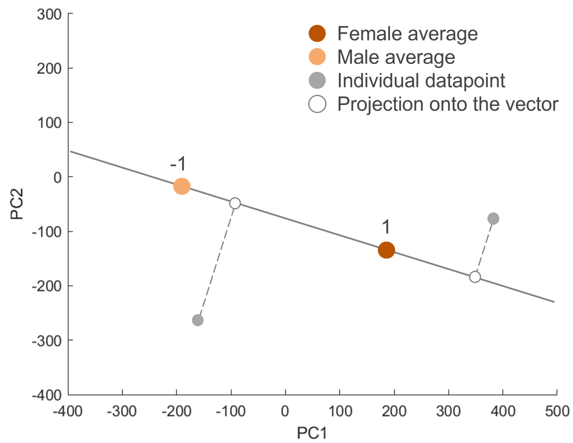

FIGURE 11. Visualization of the proposed masculinity/femininity score, illustrated in a two-dimensional space.

the most variation between male and female faces.

For each individual face, or datapoint, the position within the n-dimensional space can be projected onto the vector that runs through the mean male shape to the mean female shape. The position of the mean female shape on the vector is denoted with a value of 1, and the position of the mean male shape is denoted with a value of -1. Each projection onto the vector can be translated to a score using the distance between the point of projection on the vector and the mean male and female facial shapes. Masculinity/femininity scores with values higher than 1 represent faces that are more feminine than the average female, and values lower then -1 represent faces that are more masculine than the average male. A masculinity/femininity score of around 0 represents a facial shape that lies between the mean male and mean female shape and can be considered as neither masculine or feminine. Figure 11 shows a visualization of the proposed method to determine the masculinity/femininity score within a 2-dimensional space. The described method can be repeated for each facial region separately, using the PC scores specific for that facial region. This makes it possible to calculate masculinity/femininity scores for the entire face, as well as for the forehead, nose, chin and zygoma region. The calculation of the masculinity/femininity scores is performed using MATLAB R2022a (The MathWorks Inc., Natick, USA).

4.2.3 Scoring the faces of cisgender men and women

The masculinity/femininity score is evaluated for the cisgender dataset using the leave-one-out (LOO) validation method. For each individual face in the dataset the masculinity/femininity score is calculated once. The PC scores of the individual face for which the score is calculated, are excluded from the dataset that is used to calculate the mean male PC scores, mean female PC scores and the masculinity/femininity vector. This allows for an unbiased estimate of the masculinity/femininity score for cisgender male and female faces. Based on the masculinity/femininity score all faces in the cisgender dataset are classified as either male (masculinity/femininity score below 0) or female (masculinity/femininity score above 0). By comparing the actual sex of the individuals with the result of the classification, the classification accuracy of the masculinity/femininity score can be determined.

As demonstrated in previous chapters, the number of included PCs resulting from a threshold of 80% explained variability, might not capture the full facial shape. Therefore, the number of PCs that will be used in calculating the masculinity/femininity score will be determined based on the number of PCs that results in the highest classification accuracy demonstrated using LOO validation. The validation is started with using only the PC scores of the first PC to calculate the masculinity/femininity scores and classification accuracy. Each iteration the PC scores of the next PC are added to be used in calculating the masculinity/femininity score. The number of PCs that eventually results in the highest classification accuracy for the cisgender faces will be used for the classification of the transgender faces. This process of LOO validation and determining the suitable amount of PCs will be repeated for each of the separate facial regions.

4.2.4 Scoring the faces of transgender women

The masculinity/femininity scores for the transgender women are calculated using the number of PCs as determined using the method described above. The scores are calculated for the preoperative and postoperative 3D meshes of the face and different facial regions. The median, minimal and maximal values are determined for the preoperative and postoperative masculinity/femininity scores, as well as the mean change in the masculinity/femininity score when comparing the postoperative score to the preoperative score. The median, minimal and maximal values are determined as opposed to the mean and SD, as the number of individuals included in the transgender datasets is relatively small. The classification accuracy is calculated as the percentage of the included 3D meshes that is classified as female.

4.3. RESULTS

4.3.1 Masculinity/femininity scores of cisgender men and women

For each facial region, the number of PCs that should be included in calculating the masculinity/femininity scores is determined based on the amount of PCs resulting in the highest classification accuracy. For the full facial shape this results in using the first 20 PCs, cumulatively explaining 87.44% of the variability in facial shape within the cisgender dataset. For the separate facial regions the number of used PCs and the cumulative explained variance is 9 PCs (92.39%) for the forehead, 11 PCs (94.27%) for the nose, 19 PCs (95.84%) for the chin and 25 PCs (97.39%) for the zygoma region. The mean and SD of the masculinity/femininity scores of cisgender men and cisgender women is provided in table 18. This table also shows the classification accuracy of the masculinity/femininity score.

TABLE 18. Results of the masculinity/femininity scores for the cisgender dataset.

Facial region	Number of PCs	Mean (SD) scores cisgender women	Mean (SD) scores cisgender men	Accuracy
Full face	20	0.9949 (± 0.9867)	-0.9923 (± 1.0663)	82.93%
Forehead	9	0.9988 (± 1.3902)	-0.9987 (± 1.5872)	76.29%
Nose	11	0.9974 (± 2.1257)	-0.9833 (± 1.8995)	68.53%
Chin	19	0.9892 (± 1.7323)	-0.9825 (± 1.8707)	71.20%
Zygoma	25	0.9987 (± 1.1599)	-0.9980 (± 1.2273)	80.58%

4.3.2 Masculinity/femininity scores of transgender women

In table 19 the results of the masculinity/femininity scores for the transgender dataset are shown. The median, minimal and maximal scores are given for each of the facial regions, as well as the mean change in the masculinity/femininity score from preoperative to postoperative. The classification accuracy represents the percentage of the included faces and facial regions that is classified as female. The calculated masculinity/femininity scores for the preoperative and postoperative 3D meshes of the individual patients for the forehead, nose, chin and zygoma region are shown in respectively figure 12, 13, 14 and 15.

TABLE 19. Results of the masculinity/femininity score for the transgender dataset.

Facial region	N	Median (min, max) preoperative	Median (min, max) postoperative	Mean change	Accuracy preoperative	Accuracy postoperative
Full face	28	-0.86 (-3.12, 1.22)	1.23 (-1.61, 3.70)	+1.89	14.29%	82.14%
Forehead	22	-1.76 (-4.15, 2.05)	2.59 (-0.90, 5.56)	+3.59	27.27%	90.91%
Nose	25	-1.00 (-5.65, 2.99)	2.69 (-1.94, 7.16)	+3.76	40.00%	88.00%
Chin	13	-1.41 (-4.35, 2.41)	0.32 (-2.55, 2.13)	+1.02	46.15%	53.85%
Zygoma	8	-1.92 (-2.60, 1.25)	-0.27 (-1.35, 3.34)	+1.93	12.50%	37.50%

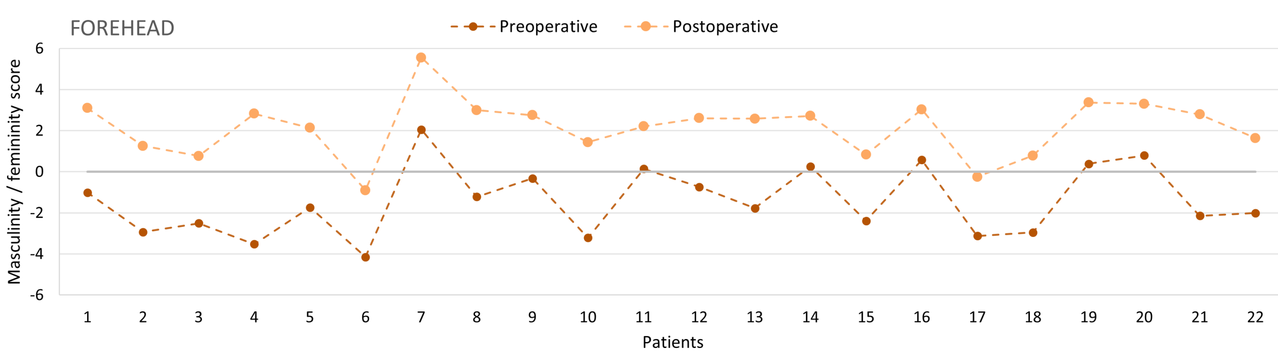

FIGURE 12. Femininity / masculinity scores for the preoperative and postoperative 3D meshes of the forehead region.

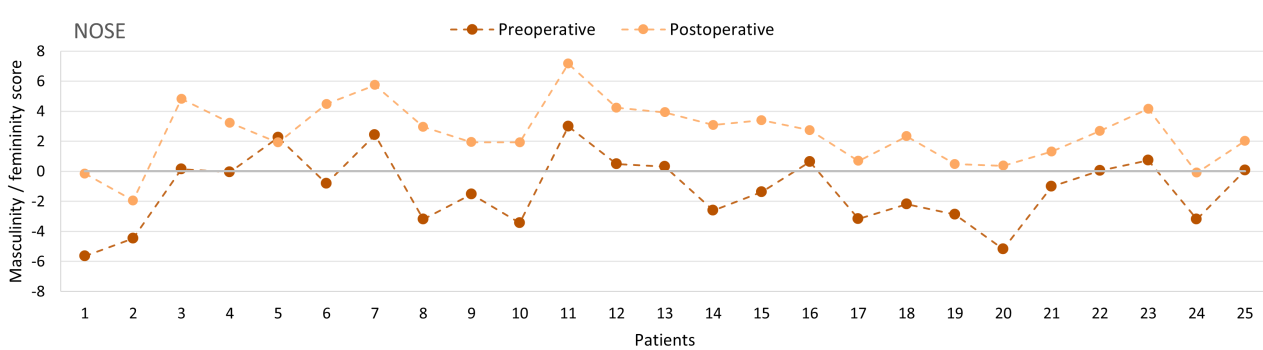

FIGURE 13. Femininity / masculinity scores for the preoperative and postoperative 3D meshes of the nose region.

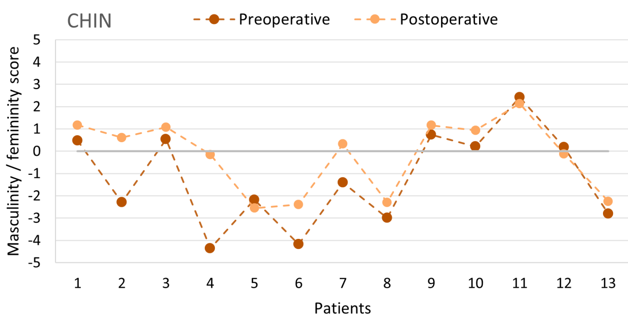

FIGURE 14. Femininity / masculinity scores for the preoperative and postoperative 3D meshes of the chin region.

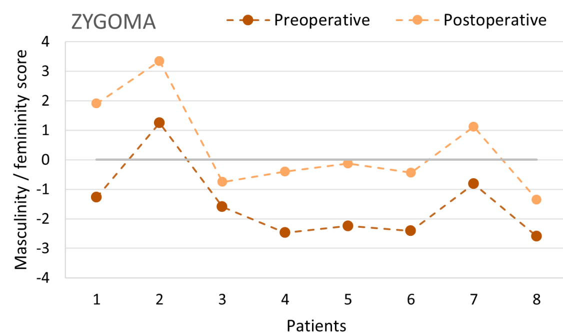

FIGURE 15. Femininity / masculinity scores for the preoperative and postoperative 3D meshes of the zygoma region.

4.4. DISCUSSION

This chapter described the development and evaluation of a masculinity/femininity score to determine the degree of masculinity or femininity in the face. The masculinity/femininity scores are determined for the preoperative and postoperative scans of transgender women for the full face, forehead, nose, chin and zygoma region.

4.4.1 Interpretation of results

The number of PCs that is included in the calculation of the masculinity/femininity score for the entire face, forehead, nose, chin and zygoma region is respectively 20, 9, 11, 19 and 25 PCs. The cumulative percentage of shape variability that is explained by these PCs ranges between 87.44% and 97.39%. This suggests that limiting the number of included PCs to 80% explained variability as used in previous chapters results in a lower classification accuracy for the masculinity/femininity score. Although the number of PCs described above resulted in the highest classification accuracies, the classification accuracy does not decrease significantly when using a lower number of PCs. Limiting the number of PCs to explain for example 90% of the shape variability will therefore not drastically influence the classification accuracy.

When looking at the results of the LOO validation of the masculinity/femininity scores of the cisgender dataset it can be observed that the mean male score and the mean female score are around -1 and 1, as determined in the definition of the masculinity/femininity score. More interesting are the SDs of the male and female distribution, as these values illustrate the amount of overlap between the male and female distributions of the masculinity/femininity scores. A higher value for the SD means that there is more overlap between the male and female distribution, which is also reflected in a lower classification accuracy. For the full facial shape the overlap between the male and female distribution of the masculinity/femininity scores is limited, with both distributions showing an SD of around 1. The corresponding classification accuracy is 82.93%. The highest SD together with the lowest classification accuracy is found for the nose area, suggesting that the masculinity/femininity score might be of limited value for differentiating between male and female noses. For the individual facial regions, the highest accuracy is observed for the zygoma region.

From the minimal and maximal values of the preoperative and postoperative masculinity/femininity scores of transgender women it can be observed that there is an overlap between the preoperative and postoperative scores. This is expected, as there was also quite some overlap present between the scores of cisgender men and cisgender women. For the forehead, nose, chin and zygoma region, it can be observed that median preoperative scores are -1 or lower, suggesting that most facial regions show a higher degree of masculinity than the mean male shape. This could be explained by the selection bias within the dataset. Only women who have undergone FFS are included within this study. It would be expected that the faces of these women show quite masculine facial features, explaining the need for FFS. As expected, the postoperative scores show higher values than the preoperative scores, illustrating that on average the face or facial region is feminised during FFS. For the forehead and nose region, it can be observed that the median postoperative masculinity/femininity score shows a value above 1, suggesting that on average the forehead and the nose show a higher degree of femininity after FFS than the mean female facial shape. For almost every individual patient, the postoperative score lies above the preoperative score, illustrating the feminization of the forehead and nose achieved by FFS.

This is not the case for the preoperative and postoperative masculinity/femininity scores of the chin, demonstrating several patients for which the masculinity/femininity score did not change that much after FFS. Although the mean observed change shows a change within the direction of the female mean, there is only a small increase in the percentage of scans that is classified as female. This could suggest that there is limited change visible in the shape of the chin, or that the change in shape is not directed along the vector explaining the difference between males and females. For the zygoma region, a similar result can be observed. There is a limited increase in the percentage of scans being classified as female after FFS. However, when looking at the preoperative and postoperative scores of the individual patients it can be observed that for each of the patients the preoperative masculinity/femininity score lies above the postoperative score. This suggests that the FFS does result in a more feminine appearance of the zygoma region, but that for a lot of the patients this change is not enough to be classified as female based on the masculinity/femininity score.

4.4.2 Study limitations

The results of the masculinity/femininity scores are based on a very small dataset of transgender women, especially for the chin and zygoma regions. Expanding the dataset will allow for more accurate and representative results for the masculinity/femininity score of the transgender faces. Within this study, for the analysis of the facial regions only transgender women who have had FFS for that specific region are included in the dataset. It would be interesting to see how the scores of these women compare to the scores of transgender women who choose not to undergo FFS for that particular facial region. This could answer the question whether the facial regions of transgender women undergoing FFS are more masculine than the facial regions of transgender women who are not undergoing FFS.

Although the proposed masculinity/femininity score shows promising results for the objective evaluation of the indications and results of FFS, further research is needed to explain the results found for the chin region. The change in shape of the chin from the preoperative to the postoperative mesh should be analyzed for the patients in which the masculinity/femininity score did not illustrate a change towards a more feminine score. This could give more insight into how the changes achieved in FFS correlate to the results of the masculinity/femininity score. It is possible that the shape of the chin is changed, but that this is not reflected within the masculinity/femininity score, as the direction of the change is not along the vector from the mean male shape to the mean female change. Another possible explanation would be that there is still too much swelling present at the time of the postoperative scan, masking the effect of FFS.

Another limitation of the proposed method to score the masculinity or femininity of the face, is that it is a purely linear approach to combining the scores of the different PCs. The masculinity/femininity score takes into account the combination of all included PCs, which proves to be a large improvement compared to the analysis of individual PCs as demonstrated in Chapter 3. However the masculinity/femininity score as proposed in this chapter is based on the projection of the PC scores of an individual point on the vector that runs from the mean male score to the mean female score. This does not take into account any possible non-linear relations between the distributions of the PC scores that could explain sexual dimorphism in the facial shape. A possible method to investigate the non-linear relations is to look into the use of an artificial intelligence (AI) model to classify the facial shape of transgender women. This possibility will be explored in the next chapter.

4.5. CONCLUSION

This chapter demonstrated the use of a masculinity/femininity score to evaluate the degree of masculinity or femininity in the faces of transgender women. The scores are calculated for the entire facial shape as well as the forehead, nose, chin and zygoma, both preoperative and postoperative. Evaluation of the masculinity/femininity score on cisgender male and female faces shows a classification accuracy of 82.93% for the entire face, with lower accuracies demonstrated for the separate facial regions. Within the transgender datasets the overall results of the masculinity/femininity score correspond with the expected feminization as a result of FFS. The classification based on the masculinity/femininity score shows an increase in the percentage of 3D meshes that is classified as female after FFS, but a significant part of the meshes is still classified as male after FFS. However, looking at the preoperative and postoperative scores of the individual transgender women it can be observed that there is an increase in the masculinity/femininity score for practically every patient for the forehead, nose and zygoma region. The chin region shows some inconclusive results, which should be further analysed. The masculinity/femininity score demonstrates promising results for the objective evaluation of the indications and results of FFS. Some further improvements and analyses are needed to make sure that the score shows accurate results for all four facial regions. Furthermore, the effect of including non-linear relations between the PCs should be evaluated using an artificial intelligence model.

Chapter 5. Gender classification of the faces of transgender women before and after facial feminization surgery using a feed-forward neural network.

5.1. INTRODUCTION

The previous chapter demonstrated the possibility of scoring the masculinity or femininity of the face based on the PC scores. The information captured in the scores of the different PCs is combined into a single masculinity/femininity score for the entire face, as well as for the forehead, nose, chin and zygoma. The masculinity/femininity score was demonstrated to be able to classify faces and facial regions with accuracies ranging between 68.53% and 82.93%. However, the masculinity/femininity score as proposed in Chapter 4 is solely based on the linear relations between the different PCs. An improvement in the classification accuracy might be achieved by taking into account the possible non-linear relations between the PCs that might play a role in the sexual dimorphism in the face.

A non-linear approach to classifying faces as male or female is with the use of an AI model. The use of AI could be of added value in the classification of faces, as AI is able to identify patterns that would be difficult to detect using traditional methods. This includes the possible non-linear relations between the different PCs and PC scores. Classification of faces as either male or female is a frequently performed task within the fields of biometrics and computer vision.[53] There is a large variety in the classification accuracies demonstrated in different studies, as the achieved accuracy largely depends on the size and homogeneity of the dataset as well as the used method. Although most studies focus on gender classification based on 2D images, several studies have used the 3D facial shape to achieve comparable classification accuracies.[54]

A feed-forward neural network (FFNN) is a type of artificial neural network that is frequently used for classification tasks and is especially suitable for high dimensional input data.[55] An FFNN consist of an input layer, multiple hidden layers and an output layer. Each layer consists of a number of nodes, which are connected to the nodes of the previous and subsequent layer. The input data moves through the different layers, with each layer processing and transforming the data to eventually create a binary classification result. As it is a feed-forward network the data only advances from the input layer towards the output layer, without feedback connections or loops embedded in the network. During training, the weights and biases of the nodes are adjusted using backpropogation and gradient descent, to minimise the outcome of the loss function. This way the network learns to recognise patterns within the input data, and can subsequently be used to classify unseen data.

In the context of FFS, an FFNN can be used to evaluate whether the faces of transgender women will be classified as male or female. This could offer more information about whether the shape of the face, or the shape of a specific facial region, is perceived as either male or female which could aid in determining the indications for FFS. Furthermore, it is possible to classify the faces of transgender women both preoperative and postoperative, to see whether the FFS has achieved the intended result of creating a face (or facial region) that is perceived as female.

The aim of this chapter is to train an FFNN to classify faces as either male or female, based on the provided PC scores. The network is trained and evaluated using the PC scores of a dataset of cisgender male and female faces. The trained network will then be applied to the PC scores describing the facial morphology of transgender women, to evaluate whether these faces and facial regions are classified as either male or female before and after FFS.

5.2. METHODS

5.2.1 Dataset

The data used in this study comprises of the same datasets as described in chapter 4. The cisgender dataset contains the PC scores of the 1236 facial 3D surface scans of cisgender males and females. This dataset is used to train the classification networks to accurately classify faces as either male or female. The transgender dataset contains the PC scores of the 28 preoperative and postoperative 3D surface scans of transgender women. This dataset is used to evaluate whether the faces and facial regions of transgender women are classified as male or female.

For the classification of the full face, the PC scores of all 28 preoperative and postoperative scans are included. However, as described in Chapter 3, not all transgender women included in the dataset have undergone the same FFS procedures. The classification for the separate facial regions is only carried out for the individuals that have undergone an FFS procedure that addresses that specific facial region. For the forehead, nose, chin and zygoma region this means that the PC scores of respectively 22, 25, 13 and 8 of the preoperative and postoperative scans are included. Furthermore, for each facial region a different number of PCs is used to describe the facial morphology. As illustrated in the previous chapters, a threshold of 80% explained variability might be too low to capture the full facial shape in PCs. Therefore, the number of PCs that is used as input for the FFNN of each of the facial regions is based on the results of Chapter 4. For the full face, forehead, nose, chin and zygoma region, this results in respectively 20, 9, 11, 19 and 25 PCs to be used as input for the FFNN.

5.2.2 The feed-forward neural network

The FFNN consists of an input layer, multiple hidden layers and an output layer. The input layer takes the PC scores of the 3D surface scans as input. The number of input nodes depends on the number of PCs that are included for the specific facial region that the model is being trained for. Gaussian noise is added to the input to reduce overfitting and creating a more robust network. Next, there are three dense fully-connected hidden layers with respectively 64, 32 and 8 nodes. Each dense layer uses a leaky rectified linear unit (ReLU) activation function, a batch normalization function and a dropout function. The output layer consists of one single neuron that uses a sigmoid activation function to output a probability score between 0 and 1 that can be uses for binary classification. The model is compiled using the Adaptive Moment

Estimation (Adam) optimizer with a learning rate of 0.001 and a binary cross-entropy loss function. The training is run for 500 epochs with a batch size of 32. The model is run on Python 3.8, using the Keras programming interface with TensorFlow 2.4 backend.

5.2.3 Evaluation of the performance of the model

The performance of the FFNN is evaluated using a 5-fold cross-validation. The dataset is partitioned into five equal folds. This means that each fold contains the PC scores of about 247 3D surface scans of cisgender male or female faces. The FFNN is trained using the data from four of the five folds. The fifth fold will be used as a validation set to evaluate the models performance. This process is repeated five times, with each fold being used as the validation set once. The average of the results of the five folds will give an estimation of the overall performance of the FFNN. The evaluation metrics are the the classification accuracy, the area under the ROC curve (AUC), the Brier score and the F1 score.

For each facial region as well as for the entire facial shape, a separate FFNN is trained and evaluated following the methods described above. After evaluating the models performance using 5-fold cross-validation, a final model will be trained on the full dataset. This results in a total of five fully trained models, which can be used to classify the entire facial shape, as well as the forehead, nose, chin and zygoma region.

5.2.4 Classification of transgender faces

The trained models can be used to classify the faces and facial regions of transgender women as either male or female. As input for the classification models, the different PC scores of the face and facial regions are used. Classification is carried out for both the preoperative and postoperative 3D surface scans by offering the PC scores as input to the different FFNNs. Evaluation is based on the percentage of the faces or facial regions that is classified as either male or as female.

5.3. RESULTS

The performance of the FFNNs in terms of classifying between cisgender male and female faces is evaluated using a 5-fold cross-validation. This is done for classification model for the entire face as well as for the forehead, nose, chin and zygoma region. The results of the individual folds of the 5-fold cross-validation are shown in Appendix G. Table 20 shows the mean evaluation metrics of the five folds for the different facial regions. The results of the classification of the preoperative and postoperative faces and facial regions of transgender women is shown in table 21.

TABLE 20. Results of the 5-fold cross-validation for the feed-forward neural networks of the different facial regions.

Facial region	Accuracy	AUC	Brier score	F1
Entire face	92.96%	0.974	0.070	0.939
Forehead	86.97%	0.926	0.130	0.891
Nose	80.18%	0.890	0.198	0.834
Chin	88.75%	0.945	0.112	0.903
Zygoma	88.59%	0.949	0.114	0.903

TABLE 21. Preoperative and postoperative classification results for the different facial regions.

		Preoperative classification				Postoperative classification			
Facial region	N	Male		Female		Male		Female	
Full face	28	19	(67.86%)	9	(32.14%)	2	(7.14%)	26	(92.86%)
Forehead	22	15	(68.18%)	7	(31.82%)	0	(0.00%)	22	(100.00%)
Nose	25	13	(52.00%)	12	(48.00%)	1	(4.00%)	24	(96.00%)
Chin	13	10	(76.92%)	3	(23.08%)	2	(15.38%)	11	(84.62%)
Zygoma	8	7	(87.50%)	1	(12.50%)	1	(12.50%)	7	(87.50%)

5.4. DISCUSSION

This study focused on developing and applying neural networks for the purpose of classifying faces as male or female. Classification is based on the PC scores that represent the morphology of the face or separate facial regions. The PC scores of the 3D meshes of the face and facial regions of transgender women are used to classify and evaluate the facial morphology before and after FFS.

5.4.1 Interpretation of the results

Looking at the overall results of the 5-fold cross-validation it can be concluded that the model performs quite well for the classification of the entire facial shape. For 92.96% of the individuals in the cisgender dataset, the classification based on the entire facial shape was correct. The classification accuracy decreases for the individual regions. This is expected as each individual region only contains a part of the information that is described by the full facial shape. For both the FFNN and the masculinity/femininity score as presented in Chapter 4, the lowest classification accuracy is found for the nose area. This implies that there is a less clear distinction in shape between cisgender male and female noses, then there is between the male and female forehead, chin and zygoma.

The FFNNs display a better classification accuracy then the masculinity/femininity score presented in Chapter 4. Both classification methods use the same PCs and PC scores as input for the different facial regions. The higher accuracy found for the AI model is likely to be contributed to the ability of the FFNN to learn and model complex, non-linear relations between the different PCs. These relations are not captured using the linear approach as presented in the masculinity/femininity score.

When looking at the classifications for transgender women, it can be observed that prior to FFS a the majority of the faces and facial regions are classified as male. This confirms the indication for FFS for the faces and facial regions for these particular transgender women. However, a lot of the 3D meshes are also already classified as female prior to FFS. This is especially the case with the nose region, where almost half of the 3D meshes are already classified as female. Even though these classifications suggest that the face or facial regions already represents as feminine, FFS has still been performed for all these women. This can be explained, as the indications of FFS are not solely based on facial shape, but also on the dysphoric feelings of the patient in respect to different aspects of the face.

For the postoperative classifications it can be observed that almost all 3D meshes of the face and separate facial regions are now classified as female. For each of the facial regions, the percentage of 3D meshes that is classified as female is either higher or quite similar to the accuracy found for the cisgender dataset. There are some individual cases that are still being classified as male after FFS. It is possible that these faces or facial regions did become a lot more feminine as a result of FFS, but are not yet feminine enough to be classified as female. However, this can not be demonstrated within this study as the FFNN does not offer information about the masculinity or femininity of the face apart from the classification.

5.4.2 Study limitations

The aim of the study presented in this chapter was to explore the possibilities of using an artificial neural network to classify the faces of transgender women. Therefore, the presented FFNN is a quite simple neural network, for which the classification accuracy might be further improved by tuning the hyperparameters or by changing the architecture of the network. The accuracy of the model can also be improved by increasing the amount of data for training the model or by providing additional information about factors that can be of influence on the facial shape. The classification could, for example, be improved by adding the age or BMI of the individual to the input of the neural network.

As described in previous chapters, being able to evaluate or classify separate facial regions is important within the context of FFS. However, the overall masculinity or femininity of the face is determined by the combination of all facial regions and the harmony between these facial features. This is something that has not been taken into account within this study. The classification has been performed for the full facial shape of transgender women before and after FFS, but these women have undergone different combinations of FFS procedures. The limited number of preoperative and postoperative scans included in this study, together with the variety in performed FFS procedures, makes it difficult to evaluate the effect of the separate FFS procedures on the classification of the full facial shape. It would be interesting to see which facial region, or which specific FFS procedure, has the most influence on the overall classification of the facial shape. This will likely be possible in future research, as more data becomes available.

Within this study it could be observed that some of the faces or facial regions of transgender women are already classified as female before FFS, or are still classified as male after FFS. Although it is expected that the facial shape did become more feminine as a result of FFS, this can not be demonstrated by solely looking at the classification of the faces. It would therefore be useful to combine the results of the classification performed by the FFNN with the results of the masculinity/femininity score as presented in Chapter 4. Another possibility would be to look at the probability scores that are used to classify the faces and facial regions. The probability score is a value between 0 and 1 that represents how confident the FFNN is in its prediction. The closer the probability score is to 1, the more confident the model is that the input should be classified as female, and the closer the score is to 0, the more confident the model is that the input scan should not be classified as female. Although the probability score is not a direct measure for the masculinity or femininity of the face, it might offer some more insight into the masculinity or femininity of the face in addition to the classification results. Analysis of the probability scores should be considered in further research.

5.5. CONCLUSION

This study illustrated the use of an FFNN to classify the faces of transgender women as either male or female, based on the PC scores describing the facial shape. A simple FFNN is trained and evaluated using a large dataset of PC scores of cisgender male and female faces, showing a classification accuracy of 92.96%. Separate networks are trained for the forehead, nose, chin and zygoma region, illustration classification accuracies of respectively 86.97%, 80.18%, 88.75% and 88.59%. Classification of the shape of the faces and facial regions of transgender women is performed for both the preoperative and postoperative 3D meshes. Preoperative the majority of the faces and facial regions is classified as male, while postoperative almost all faces and facial regions are classified as female. This illustrates both the need for FFS, as well as the successful results of FFS in terms of creating a feminine shape of the face and facial regions. Combining the classification results of the FFNN with the masculinity/femininity score as presented in Chapter 4 could offer more information about the change in terms of masculinity and femininity of the face that is achieved during FFS. This is especially useful for the faces and facial regions that are already classified as female before FFS, or for the faces and facial regions that are still classified as male after FFS. The classification of the face and different facial regions offers an objective evaluation of the facial shape and could be of added value in determining which specific FFS procedures to perform.

General discussion & conclusion

The research presented in this thesis focused on the use of facial 3D surface scans to analyse the facial morphology of transgender women in comparison to cisgender male and female faces. A PCA is performed on the coordinates of the 3D surface scans, and the PC scores are used to evaluate differences between cisgender male, cisgender female and transgender female faces. The analyses are performed for the entire facial shape, as well as for the forehead, nose, chin and zygoma region. The found differences between cisgender male and female faces largely correspond to the facial features that explain sexual dimorphism of the face as described in literature. When comparing the facial shape of transgender women prior to FFS, it could be observed that the facial shape shows a large correspondence to the facial shape of cisgender men. However, there are also differences observed, which might indicate the effects of gender-affirmative hormone treatment on the facial shape of transgender women. The effect of different FFS procedures on the facial shape is evaluated by comparing preoperative and postoperative facial 3D surface scans of transgender women. A method for scoring the masculinity and femininity of the facial shape is proposed, as well as an AI model to classify the faces of transgender women as either male or female. The observed changes in the scored masculinity or femininity and the classification of the faces and facial regions correspond with the aim of FFS to change the more masculine facial features into more feminine facial features.

An important consideration to be made is how the proposed facial shape analysis, masculinity/femininity scores and classifications could be implemented in clinical care. As the main goal of FFS is to reduce gender dysphoria, the indications for FFS should always be focused on both the facial shape as well as the dysphoric feelings of the patient in regards to the face. Facial shape analysis and classification of the different facial regions can be used as additional information when determining which FFS procedures should be performed for the individual patient. Furthermore, being able to 'prove' that the face or facial regions of a transgender women are in fact classified as male might help in providing insurance coverage for the FFS procedures. It provides an objective approach to compare the facial shape of transgender women to the facial shape of cisgender women. However, for some of the transgender women the faces or facial regions are already classified as female before undergoing FFS. It can be questioned whether performing these procedures can be justified when solely looking at the facial shape and classification results. When an objective evaluation, either in the form of a classification result or a masculinity/femininity score, should become part of for example the request for insurance coverage, it has to be considered that this would not be beneficial for all women. For transgender women whose faces or specific facial regions are already classified as female prior to FFS, using an objective evaluation might complicate or limit the coverage by healthcare insurance.

Future research should be focused on reducing the study limitations and improving the proposed methods as described in the individual chapters. Some important limitations that should be overcome in future research include the relatively small number of available postoperative 3D surface scans together with the large variety of performed FFS procedures. As more patients are being operated on and more data becomes available, this problem will decrease in future research. Another limitation includes the limited ability to generalize the results of the study to individuals of different ethnicities. Further research should be focused on repeating the proposed methods using datasets of individuals of different ethnicities, as a large part of the patient population for FFS consists of non-Caucasion individuals. The same applies for the age limit used in this study. In future research, the proposed methods can be evaluated for the facial shape of individuals aged above 50 years, so that the facial shape analysis can eventually be applied to a larger part of the patient population. Furthermore, there is limited research available about the influence of gender-affirmative hormone treatment on the facial shape of transgender women. Within the context of FFS it would be interesting to see which changes in facial shape are induced by the hormones and whether these changes influence the indications and results of FFS.

The study presented in this thesis demonstrated different approaches to evaluate the facial shape of transgender women with the use of 3D surface scans. This allows for an objective and quantitative analysis of facial morphology, which allow for a better understanding and quantification of the surgical indications and results of FFS. Further improvement of the proposed methods is needed before clinical implementations can be considered.

REFERENCES

[1] K.J. Zucker. Epidemiology of gender dysphoria and transgender identity. *Sexual health*, 14(5):404–411, 2017. doi: 10.1071/SH17067.

[2] W. Huijnk, R. Damen, and L. Kampen, van. LHBT-monitor 2022. De leefsituatie van lesbische, homoseksuele, biseksuele en transgender personen in Nederland. Technical report, Den Haag: Sociaal en Cultureel Planbureau., 2022.

[3] L. Kuyper and C. Wijsen. Gender identities and gender dysphoria in the Netherlands. *Archives of Sexual Behavior*, 43(2):377–385, 2014. doi: 10.1007/S10508-013-0140-Y/TABLES/5.

[4] P. Tawa, N. Brault, V. Luca-Pozner, L. Ganry, G. Chebbi, M. Atlan, and Q. Qassemyar. Three-Dimensional Custom-Made Surgical Guides in Facial Feminization Surgery: Prospective Study on Safety and Accuracy. *Aesthetic surgery journal*, 41(11):NP1368–NP1378, 10 2021. doi: 10.1093/asj/sjab032.

[5] T.A. Ainsworth and J.H. Spiegel. Quality of life of individuals with and without facial feminization surgery or gender reassignment surgery. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 19 (7):1019–1024, 9 2010. doi: 10.1007/s11136-010-9668-7.

[6] L.M. Wesp and M.B. Deutsch. Hormonal and Surgical Treatment Options for Transgender Women and Transfeminine Spectrum Persons. *Psychiatric Clinics of North America*, 40(1):99–111, 2017. doi: 10.1016/J.PSC.2016.10.006.

[7] S.A. Martin, S.D. Morrison, V Patel, F. Capitán-Cañadas, A. Sánchez-García, M. Rodríguez-Conesa, R.J. Bellinga, D. Simon, L. Capitán, T. Satterwhite, and R. Nazerali. Social Perception of Facial Feminization Surgery Outcomes: Does Gender Identity Alter Gaze? *Aesthetic surgery journal*, 41(10):1207–1215, 9 2021. doi: 10.1093/asj/sjaa377.

[8] E Coleman, A.E. Radix, W.P. Bouman, G.R. Brown, A.L.C. de Vries, M.B. Deutsch, R Ettner, L Fraser, ..., and J Arcelus. Standards of Care for the Health of Transgender and Gender Diverse People, Version 8. *International Journal of Transgender Health*, 23: 1–259, 2022. doi: 10.1080/26895269.2022.2100644.

[9] B. Colebunders, S. Brondeel, S. D'Arpa, P. Hoebeke, and S. Monstrey. An Update on the Surgical Treatment for Transgender Patients. *Sexual Medicine Reviews*, 5(1):103–109, 2017. doi: 10.1016/J.SXMR.2016.08.001.

[10] B. Safa, W.C. Lin, A.M. Salim, J.C. Deschamps-Braly, and M.M. Poh. Current Concepts in Masculinizing Gender Surgery. *Plastic and Reconstructive Surgery*, 143(4):857–871, 2019. doi: 10.1097/PRS.0000000000005453.

[11] B. Safa, W.C. Lin, A.M. Salim, J.C. Deschamps-Braly, and M.M. Poh. Current Concepts in Feminizing Gender Surgery. *Plastic and Reconstructive Surgery*, 143(5):1081–1091, 2019. doi: 10.1097/PRS.0000000000005595.

[12] K. Altman. Facial feminization surgery: current state of the art. *International journal of oral and maxillofacial surgery*, 41(8): 885–894, 8 2012. doi: 10.1016/j.ijom.2012.04.024.

[13] S.D. Morrison, K.S. Vyas, S. Motakef, K.M. Gast, M.T. Chung, V. Rashidi, T. Satterwhite, W. Kuzon, and P.S. Cederna. Facial Feminization: Systematic Review of the Literature. *Plastic and reconstructive surgery*, 137(6):1759–1770, 6 2016. doi: 10.1097/PRS.0000000000002171.

[14] S.D. Morrison, F. Capitán-Cañadas, A. Sánchez-García, D.C. Ludwig, J.P. Massie, I.T. Nolan, M. Swanson, M. Rodríguez-Conesa, J.B. Friedrich, P.S. Cederna, R.J. Bellinga, D. Simon, L. Capitán, and T. Satterwhite. Prospective Quality-of-Life Outcomes after Facial Feminization Surgery: An International Multicenter Study. *Plastic and reconstructive surgery*, 145(6):1499–1509, 6 2020. doi: 10.1097/PRS.0000000000006837.

[15] M. Somenek. Gender-Related Facial Surgical Goals. *Facial plastic surgery : FPS*, 34(5):474–479, 10 2018. doi: 10.1055/s-0038-1670648.

[16] J.C. Deschamps-Braly. Facial Gender Confirmation Surgery: Facial Feminization Surgery and Facial Masculinization Surgery. *Clinics in plastic surgery*, 45(3):323–331, 7 2018. doi: 10.1016/j.cps.2018.03.005.

[17] H.M. Juszczak, Z. Fridirici, P.D. Knott, A.M. Park, and R. Seth. An update in facial gender confirming surgery. *Current Opinion in Otolaryngology and Head and Neck Surgery*, 27(4):243–252, 2019. doi: 10.1097/MOO.0000000000000549.

[18] B.F. Chaya, Z.P. Berman, D. Boczar, N. Siringo, R. Rodriguez Colon, J. Trilles, G.K. Diep, and E.D. Rodriguez. Current Trends in Facial Feminization Surgery: An Assessment of Safety and Style. *The Journal of craniofacial surgery*, 32(7):2366–2369, 2021. doi: 10.1097/SCS.0000000000007785.

[19] A. Dubov and L. Fraenkel. Facial Feminization Surgery: The Ethics of Gatekeeping in Transgender Health. *The American journal of bioethics : AJOB*, 18(12):3–9, 12 2018. doi: 10.1080/15265161.2018.1531159.

[20] S. La Padula, B. Hersant, H. Chatel, P. Aguilar, R. Bosc, G. Roccaro, R. Ruiz, and J.P. Meningaud. One-step facial feminization surgery: The importance of a custom-made preoperative planning and patient satisfaction assessment. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 72(10):1694–1699, 2019. doi: 10.1016/J.BJPS.2019.06.014.

[21] A.R. Tirrell, A.A. Abu El Hawa, J.C. Bekeny, B.L. Chang, and G.D. Corral. Facial Feminization Surgery: A Systematic Review of Perioperative Surgical Planning and Outcomes. *Plastic and Reconstructive Surgery Global Open*, 10(3):4210, 2022. doi: 10.1097/GOX.0000000000004210.

[22] M. Fisher, S.M. Lu, K. Chen, B. Zhang, M. Di Maggio, and J.P. Bradley. Facial Feminization Surgery Changes Perception of Patient Gender. *Aesthetic Surgery Journal*, 40(7):703–709, 2020. doi: 10.1093/ASJ/SJZ303.

[23] P. Mitteroecker and P. Gunz. Advances in Geometric morphometrics. *Evolutionary Biology*, 36(2):235–247, 2009. doi: 10.1007/S11692-009-9055-X/FIGURES/7.

[24] F. Ambellan, H. Lamecker, C. von Tycowicz, and S. Zachow. Statistical Shape Models: Understanding and Mastering Variation in Anatomy. *Advances in experimental medicine and biology*, 1156:67–84, 2019. doi: 10.1007/978-3-030-19385-0{_}5.

[25] A. Samal, V. Subramani, and D. Marx. Analysis of sexual dimorphism in human face. *Journal of Visual Communication and Image Representation*, 18(6):453–463, 2007. doi: 10.1016/J.JVCIR.2007.04.010.

[26] J.M. Robertson, B.E. Kingsley, and G.C. Ford. Sexually Dimorphic Faciometrics in Humans From Early Adulthood to Late Middle Age: Dynamic, Declining, and Differentiated. *Evolutionary Psychology*, 15(3), 2017. doi: 10.1177/1474704917730640/ASSET/IMAGES/LARGE/10.1177{_}1474704917730640-FIG1.JPEG.

[27] J.J. Bannister, H. Juszczak, J.D. Aponte, D.C. Katz, P.D. Knott, S.M. Weinberg, B. Hallgrímsson, N.D. Forkert, and R. Seth. Sex Differences in Adult Facial Three-Dimensional Morphology: Application to Gender-Affirming Facial Surgery. *Facial Plastic Surgery & Aesthetic Medicine*, 24(S2), 2022. doi: 10.1089/FPSAM.2021.0301.

[28] O. Catz and M.B. Lewis. Exploring distinctiveness, attractiveness and sexual dimorphism in actualized face-spaces. *Visual Cognition*, 28(9):453–469, 2020. doi: 10.1080/13506285.2020.1797967.

[29] H.S. Matthews, A.J. Penington, R. Hardiman, Y. Fan, J.G. Clement, N.M. Kilpatrick, and P.D. Claes. Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Scientific Reports*, 8(1):4771, 2018. doi: 10.1038/s41598-018-22752-5.

[30] M.J. Kesterke, Z.D. Raffensperger, C.L. Heike, M.L. Cunningham, J.T. Hecht, C.H. Kau, N.L. Nidey, L.M. Moreno, G.L. Wehby, M.L. Marazita, and S.M. Weinberg. Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. *Biology of Sex Differences*, 7(1):1–14, 2016. doi: 10.1186/S13293-016-0076-8/FIGURES/6.

[31] M. Tebbens, N.M. Nota, N.P.T.J. Liberton, B.A. Meijer, B.P.C. Kreukels, T. Forouzanfar, R.M. Verdaasdonk, and M. den Heijer. Gender-Affirming Hormone Treatment Induces Facial Feminization in Transwomen and Masculinization in Transmen: Quantification by 3D Scanning and Patient-Reported Outcome Measures. *The journal of sexual medicine*, 16(5):746–754, 2019. doi: 10.1016/J.JSXM.2019.02.011.

[32] N.E. Pears, C. Duncan, W.A.P. Smith, and H. Dai. The Headspace dataset. Technical report, 2018.

[33] J.D. White, A. Ortega-Castrillón, H. Matthews, A.A. Zaidi, O. Ekrami, J. Snyders, Y. Fan, T. Penington, S. Van Dongen, M.D. Shriver, and P. Claes. MeshMonk: Open-source large-scale intensive 3D phenotyping. *Scientific Reports*, 9(1), 2019. doi: 10.1038/S41598-019-42533-Y.

[34] TheWebMonks. GitHub - Meshmonk: Open mesh-to-mesh registration framework. URL https://github.com/TheWebMonks/meshmonk.

[35] Z. Skomina, M. Verdenik, and N.I. Hren. Effect of aging and body characteristics on facial sexual dimorphism in the Caucasian Population. *PloS one*, 15(5), 2020. doi: 10.1371/JOURNAL.PONE.0231983.

[36] G. Rhodes. The Evolutionary Psychology of Facial Beauty. *Annual Review of Psychology*, 57:199–226, 2006. doi: 10.1146/ANNUREV.PSYCH.57.102904.190208.

[37] R.S.S. Kramer, A.L. Jones, and R. Ward. A lack of sexual dimorphism in width-to-height ratio in white European faces using 2D photographs, 3D scans, and anthropometry. *PloS one*, 7(8), 2012. doi: 10.1371/JOURNAL.PONE.0042705.

[38] R.S.S. Kramer. Sexual dimorphism of facial width-to-height ratio in human skulls and faces: A meta-analytical approach. *Evolution and Human Behavior*, 38(3):414–420, 2017. doi: 10.1016/J.EVOLHUMBEHAV.2016.12.002.

[39] S. Ritz-Timme, P. Gabriel, J. Tutkuviene, P. Poppa, Z. Obertová, D. Gibelli, D. De Angelis, M. Ratnayake, R. Rizgeliene, A. Barkus, and C. Cattaneo. Metric and morphological assessment of facial features: A study on three European populations. *Forensic Science International*, 207(1-3):1–239, 2011. doi: 10.1016/J.FORSCIINT.2011.01.035.

[40] K. Kleisner, P. Tureček, S. C. Roberts, J. Havlíček, J.V. Valentova, R.M. Akoko, J.D. Leongómez, S. Apostol, M.A.C. Varella, and S. A. Saribay. How and why patterns of sexual dimorphism in human faces vary across the world. *Scientific Reports 2021 11:1*, 11(1):1–14, 2021. doi: 10.1038/s41598-021-85402-3.

[41] C. Duncan, N.E. Pears, H. Dai, W.A.P. Smith, and P. O'higgins. Applications of 3D Photography in Craniofacial Surgery. *Journal of Pediatric Neurosciences*, 17(Suppl 1):21–28, 9 2022. doi: 10.4103/JPN.JPN{_}48{_}22.

[42] H.T. Lübbers, L. Medinger, A. Kruse, K.W. Grätz, and F. Matthews. Precision and accuracy of the 3dmd photogrammetric system in craniomaxillofacial application. *Journal of Craniofacial Surgery*, 21(3):763–767, 2010. doi: 10.1097/SCS.0B013E3181D841F7.

[43] P.G.M. Knoops, C.A.A. Beaumont, A. Borghi, N. Rodriguez-Florez, R.W.F. Breakey, W. Rodgers, F. Angullia, N.U.O. Jeelani, S. Schievano, and D.J. Dunaway. Comparison of three-dimensional scanner systems for craniomaxillofacial imaging. *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 70(4):441–449, 2017. doi: 10.1016/j.bjps.2016.12.015.

[44] Y. Fan, A. Penington, N.H. Kilpatrick, P. Schneider, J. Clement, P. Claes, and H. Matthews. Quantification of mandibular sexual dimorphism during adolescence. *Journal of Anatomy*, 234(5):709–717, 2019. doi: 10.1111/JOA.12949.

[45] D. Franklin, P. O'Higgins, C.E. Oxnard, and I. Dadour. Sexual dimorphism and population variation in the adult mandible : Forensic applications of geometric morphometrics. *Forensic science, medicine, and pathology*, 3(1):15–22, 2007. doi: 10.1385/FS MP:3:1:15.

[46] R.J. Hennessy, S. McLearie, A. Kinsella, and J.L. Waddington. Facial surface analysis by 3D laser scanning and geometric morphometrics in relation to sexual dimorphism in cerebral–craniofacial morphogenesis and cognitive function. *Journal of Anatomy*, 207(3):283–295, 2005. doi: 10.1111/J.1469-7580.2005.00444.X.

[47] Z.M. Thayer and S.D. Dobson. Sexual dimorphism in chin shape: Implications for adaptive hypotheses. *American Journal of Physical Anthropology*, 143(3):417–425, 2010. doi: 10.1002/AJPA.21330.

[48] A.G. Becking, D.B. Tuinzing, J.J. Hage, and L.J.G. Gooren. Transgender Feminization of the Facial Skeleton. *Clinics in Plastic Surgery*, 34(3):557–564, 2007. doi: 10.1016/j.cps.2007.04.009.

[49] K. De Boulle, N. Furuyama, I. Heydenrych, Keaney, A. Rivkin, V. Wong, and M. Silberberg. Considerations for the Use of Minimally Invasive Aesthetic Procedures for Facial Remodeling in Transgender Individuals. *Clinical, Cosmetic and Investigational Dermatology*, 14:513, 2021. doi: 10.2147/CCID.S304032.

[50] K.J. Pereira, C.S.A. da Silva, J. Havlíček, K. Kleisner, M.A.C. Varella, O. Pavlovic, and J.V. Valentova. Femininity-masculinity and attractiveness – Associations between self-ratings, third-party ratings and objective measures. *Personality and Individual Differences*, 147:166–171, 9 2019. doi: 10.1016/J.PAID.2019.04.033.

[51] M. Komori, S. Kawamura, and S. Ishihara. Multiple mechanisms in the perception of face gender: Effect of sex-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3):626–633, 2011. doi: 10.1037/A0020369.

[52] P. Mitteroecker, S. Windhager, G.B. Müller, and K. Schaefer. The Morphometrics of "Masculinity" in Human Faces. *PLoS ONE*, 10(2), 2015. doi: 10.1371/JOURNAL.PONE.0118374.

[53] S.A. Khan, M. Ahmad, M. Nazir, N. Riaz, S. Zulfikar, and A. Bhutto. A Comparative Analysis of Gender Classification Techniques. *International Journal of Bio-Science and Bio-Technology*, 5(4), 2013.

[54] L. Tian, J. Liu, and W. Guo. Three-Dimensional Face Reconstruction Using Multi-View-Based Bilinear Model. *Sensors*, 19(3):459, 2019. doi: 10.3390/S19030459.

[55] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 1 edition, 2016. URL https://www.deeplearningbook.org/.

Appendix

A. RESULTS OF THE PCA OF THE FACE

A.1 Variance explained by the PCs

TABLE 22. Percentage of variability explained by the first 30 principal components.

PC	%	Cum. %	PC	%	Cum. %	PC	%	Cum. %
PC1	26.21	26.21	PC11	1.71	78.71	PC21	0.54	87.98
PC2	18.14	44.35	PC12	1.64	80.35	PC22	0.52	88.50
PC3	7.83	52.18	PC13	1.26	81.61	PC23	0.49	88.99
PC4	7.16	59.34	PC14	1.07	82.68	PC24	0.46	89.45
PC5	4.37	63.71	PC15	1.01	83.69	PC25	0.39	89.84
PC6	3.83	67.55	PC16	0.90	84.60	PC26	0.38	90.21
PC7	2.80	70.35	PC17	0.87	85.46	PC27	0.36	90.57
PC8	2.54	72.90	PC18	0.80	86.26	PC28	0.34	90.91
PC9	2.24	75.14	PC19	0.62	86.89	PC29	0.31	91.22
PC10	1.86	77.00	PC20	0.56	87.44	PC30	0.28	91.50

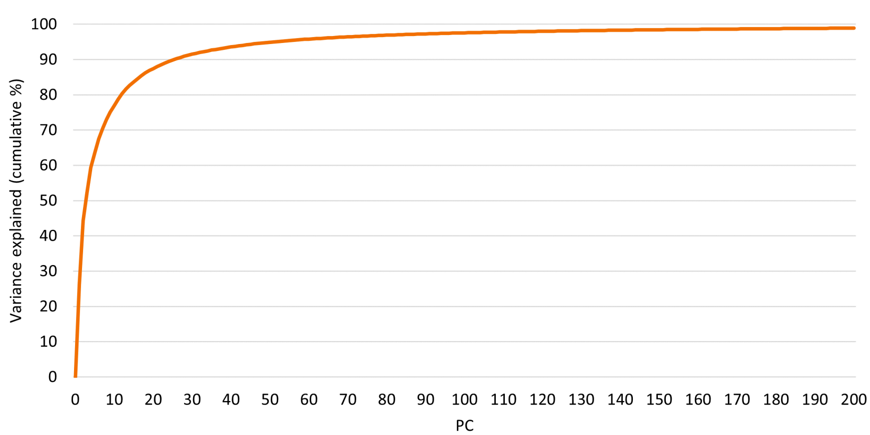

FIGURE 16. Cumulative percentage of variability explained by the first 100 principal components.

A.2 Visualizations of the facial shape variability

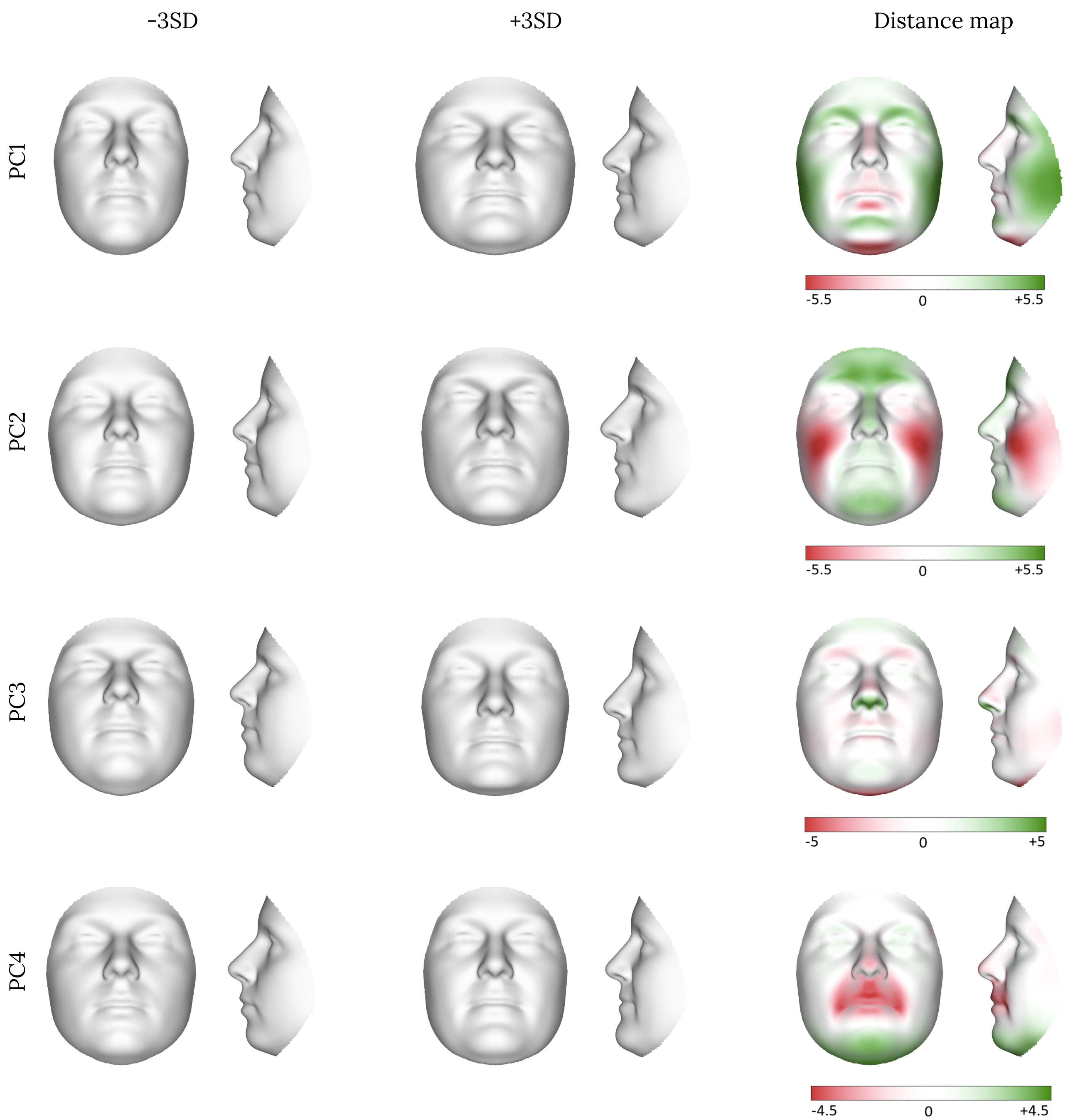

FIGURE 17. (Continued) Visualization of the shape variability captured in the first twelve principal components. The first and second column show the frontal and profile views of the reconstructed faces that differ three standard deviations from the mean face in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed faces. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

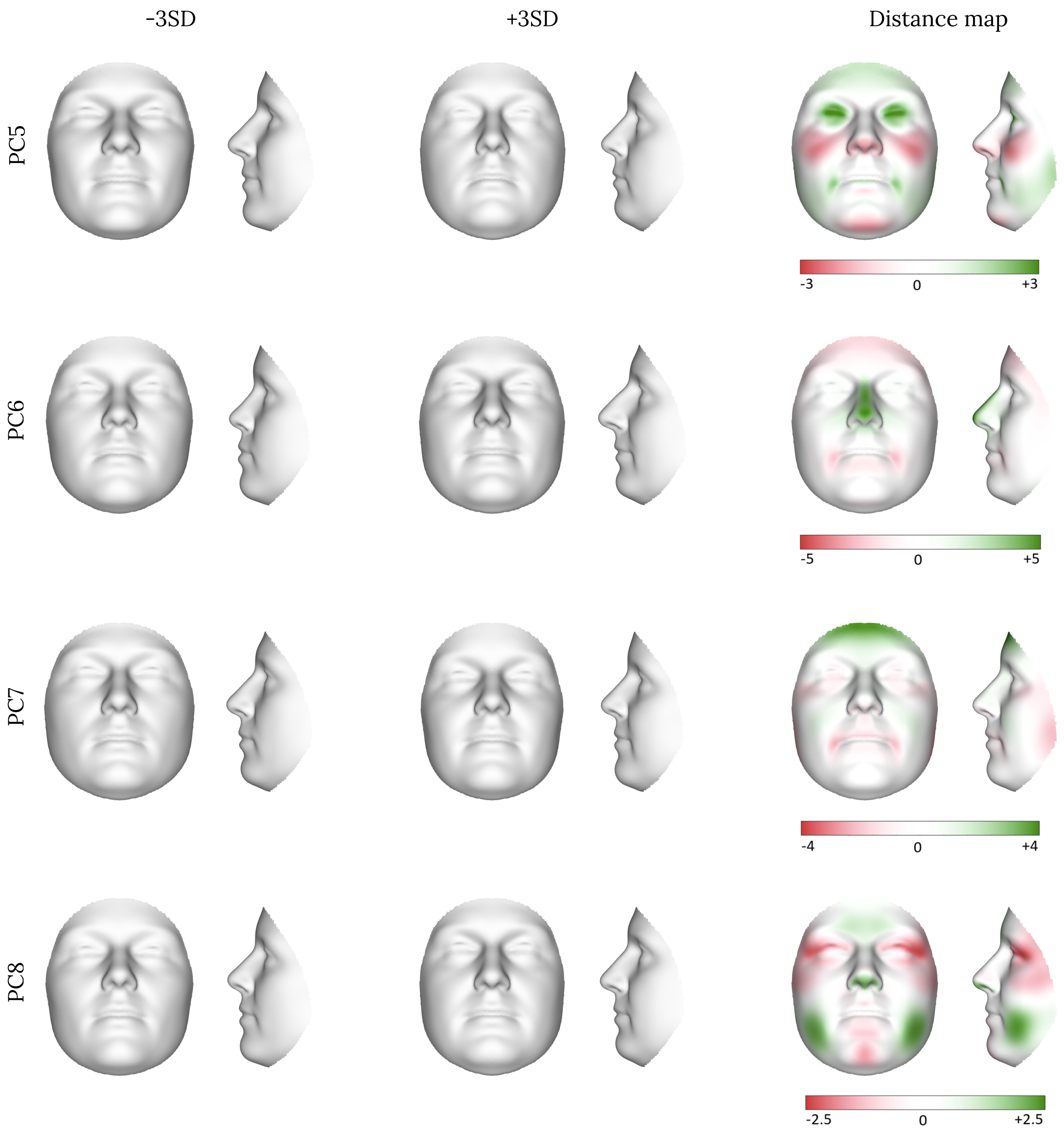

FIGURE 17. (Continued) Visualization of the shape variability captured in the first twelve principal components. The first and second column show the frontal and profile views of the reconstructed faces that differ three standard deviations from the mean face in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed faces. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

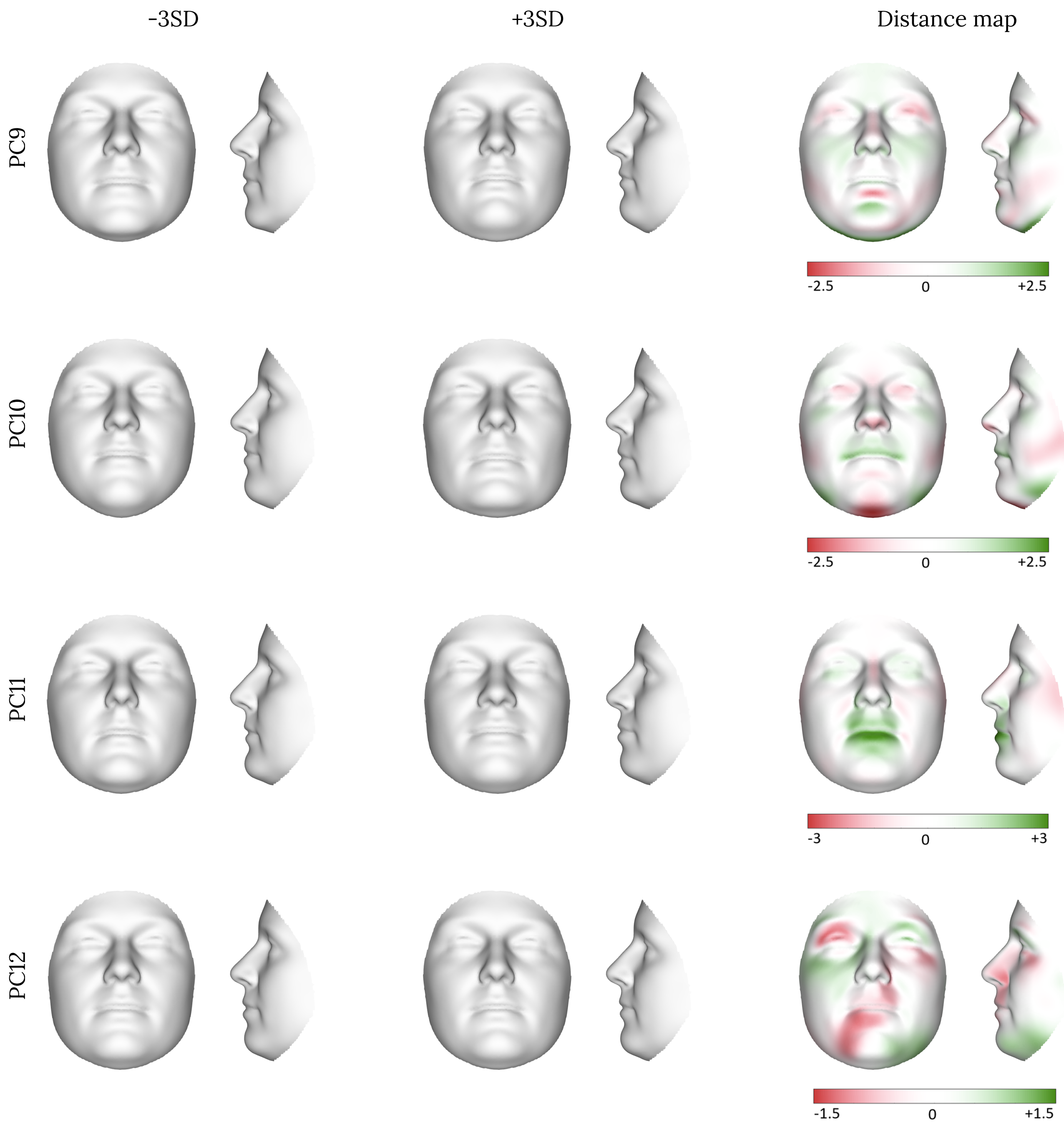

FIGURE 17. (Continued) Visualization of the shape variability captured in the first twelve principal components. The first and second column show the frontal and profile views of the reconstructed faces that differ three standard deviations from the mean face in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed faces. The distances are measured in mm the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

A.3 Descriptive statistics

TABLE 23. Descriptive statistics for the cisgender male, cisgender female and transgender female group for the principal component scores for the first twelve principal components.

PC	Group	N	Mean	SD
PC1	Cisgender male	518	-8.085	132.503
	Cisgender female	718	4.754	121.534
	Transgender female	203	14.366	134.542
PC2	Cisgender male	518	77.092	85.561
	Cisgender female	718	-55.336	80.594
	Transgender female	203	49.471	91.094
PC3	Cisgender male	518	-22.076	66.857
	Cisgender female	718	16.050	66.775
	Transgender female	203	-32.826	66.956
PC4	Cisgender male	518	-1.688	67.599
	Cisgender female	718	.611	64.764
	Transgender female	203	-27.102	64.849
PC5	Cisgender male	518	3.377	52.226
	Cisgender female	718	-1.833	50.416
	Transgender female	203	-5.527	46.951
PC6	Cisgender male	518	1.864	48.463
	Cisgender female	718	-1.476	48.588
	Transgender female	203	4.131	52.647
PC7	Cisgender male	518	-9.523	40.505
	Cisgender female	718	6.694	40.775
	Transgender female	203	-10.454	41.532
PC8	Cisgender male	518	12.327	39.789
	Cisgender female	718	-9.243	36.089
	Transgender female	203	-5.105	39.301
PC9	Cisgender male	518	-5.429	39.218
	Cisgender female	718	4.247	34.366
	Transgender female	203	22.475	39.289
PC10	Cisgender male	518	-3.894	34.619
	Cisgender female	718	2.898	32.828
	Transgender female	203	3.775	33.758
PC11	Cisgender male	518	2.136	31.585
	Cisgender female	718	-1.439	32.965
	Transgender female	203	-.540	31.161
PC12	Cisgender male	518	-1.602	32.232
	Cisgender female	718	1.383	31.211
	Transgender female	203	.854	33.365

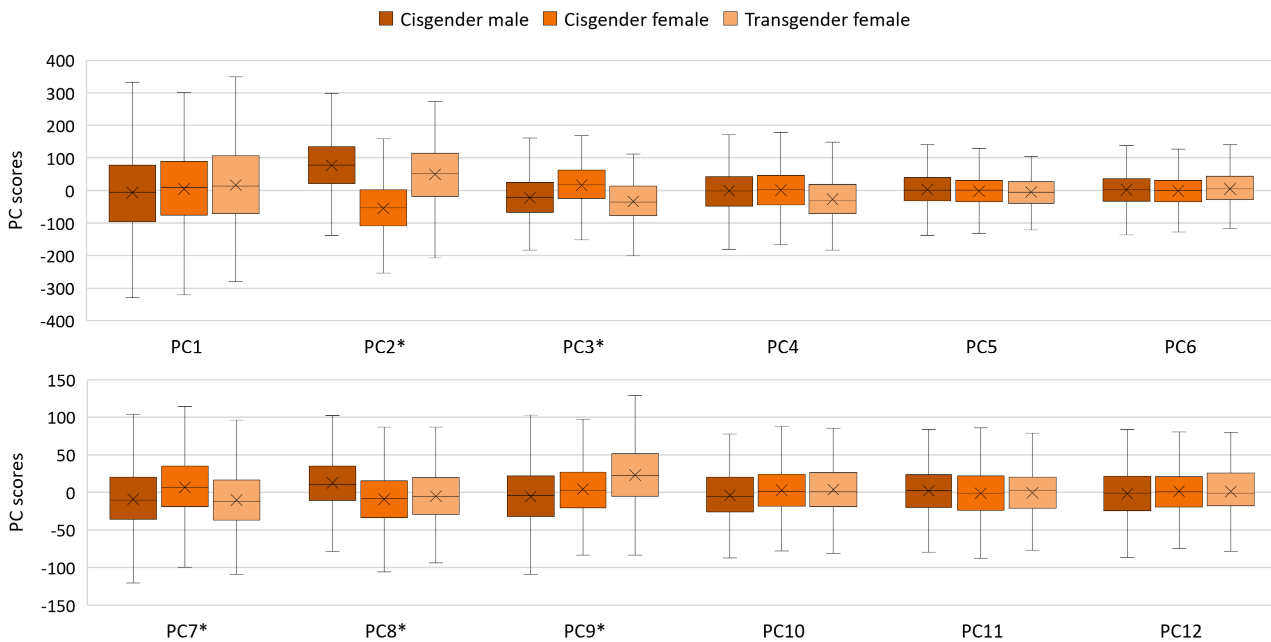

FIGURE 18. Distribution of the principal component scores for cisgender male, cisgender female and transgender female faces for the first twelve principal components. *The mean differences among groups is significant at the .05 level, with an effect size larger then 0.20.

B. RESULTS OF THE PCA FOR SEPARATE FACIAL REGIONS

B.1 Variance explained by the PCs

TABLE 24. Percentage of variability explained by the first twelve principal components for the four facial regions.

	Forehead		Nose		Chin		Zygoma	
PC	**%**	**Cum. %**	**%**	**Cum. %**	**%**	**Cum. %**	**%**	**Cum. %**
PC1	50.17	50.17	40.64	40.64	43.47	43.47	43.58	43.58
PC2	15.32	65.48	31.35	71.99	20.94	64.41	21.98	65.55
PC3	9.56	75.05	5.39	77.38	6.47	70.88	9.05	74.60
PC4	6.12	81.17	4.79	82.17	5.67	76.55	4.56	79.16
PC5	3.68	84.85	3.92	86.09	3.62	80.16	3.29	82.45
PC6	3.10	87.95	2.99	89.08	2.77	82.93	2.57	85.02
PC7	2.06	90.01	1.66	90.74	2.27	85.20	2.27	87.29
PC8	1.28	91.29	1.22	91.96	2.06	87.26	2.06	89.35
PC9	1.10	92.39	0.93	92.89	1.76	89.02	1.45	90.79
PC10	0.98	93.38	0.74	93.63	1.38	90.39	1.10	91.89
PC11	0.78	94.16	0.64	94.27	1.21	91.60	0.71	92.60
PC12	0.70	94.86	0.51	94.78	0.92	92.52	0.70	93.30

C. RESULTS OF THE PCA FOR THE FOREHEAD REGION

C.1 Visualizations of the shape variability of the forehead region

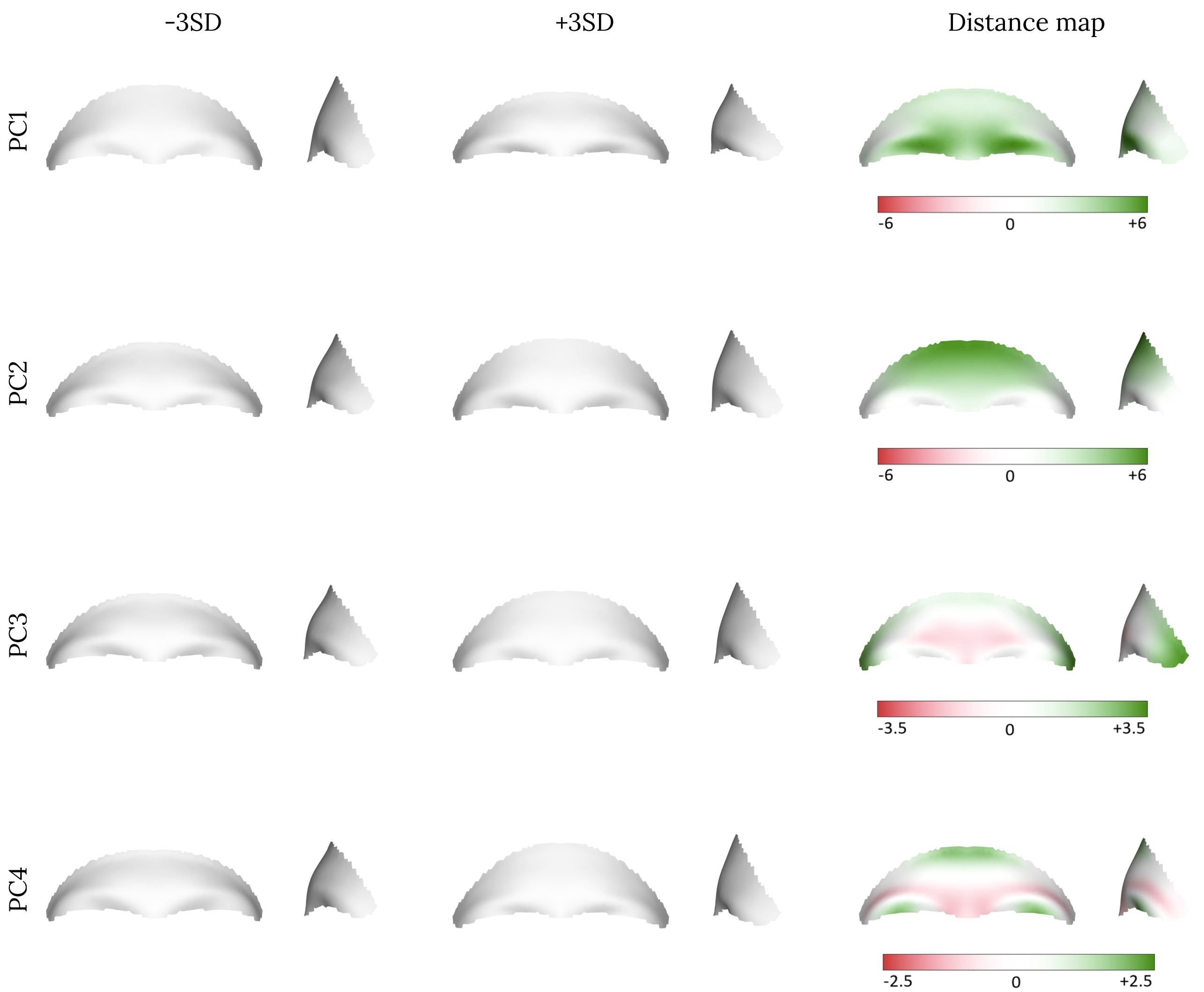

FIGURE 19. Visualization of the shape variability of the forehead captured in the first four principal components. The first and second column show the frontal and profile views of the reconstructed forehead regions that differ three standard deviations from the mean shape in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed regions. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

C.2 Descriptive statistics

TABLE 25. Descriptive statistics for the cisgender male, cisgender female and transgender female group for the principal component scores of the first four principal components.

PC	Group	N	Mean	SD
PC1	Cisgender male	518	39.173	69.131
	Cisgender female	718	-28.261	61.404
	Transgender female	203	42.266	73.653
PC2	Cisgender male	518	2.912	38.315
	Cisgender female	718	-2.101	41.433
	Transgender female	203	-6.454	41.223
PC3	Cisgender male	518	-18.138	27.213
	Cisgender female	718	13.086	28.213
	Transgender female	203	-4.757	28.145
PC4	Cisgender male	518	-4.653	27.448
	Cisgender female	718	3.357	23.311
	Transgender female	203	-3.033	22.930

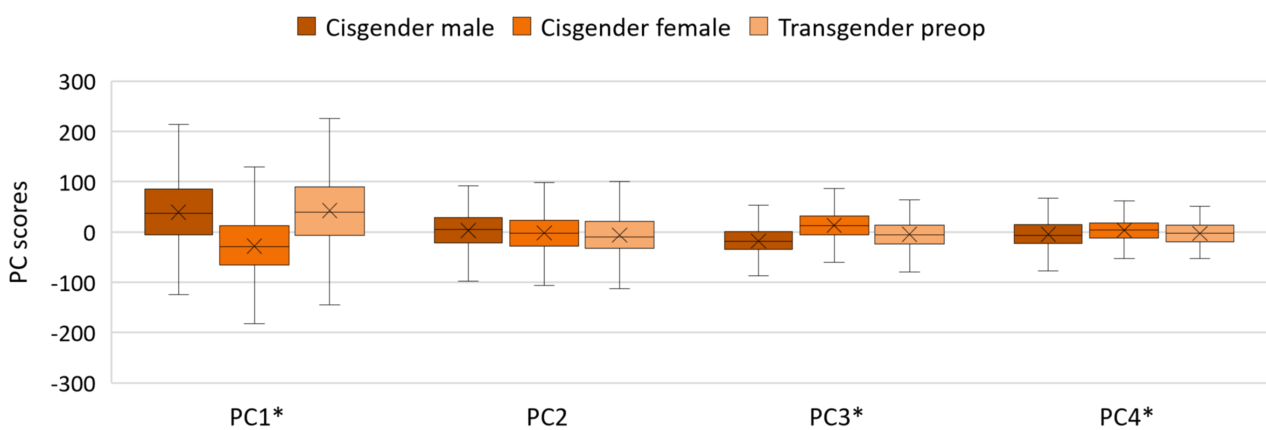

FIGURE 20. Distribution of the principal component scores for the cisgender male, cisgender female and transgender female forehead region for the first 4 principal components. *The mean differences among groups is significant at the .05 level, with an effect size larger then 0.20.

D. RESULTS OF THE PCA FOR THE NOSE REGION

D.1 Visualizations of the shape variability of the nose region

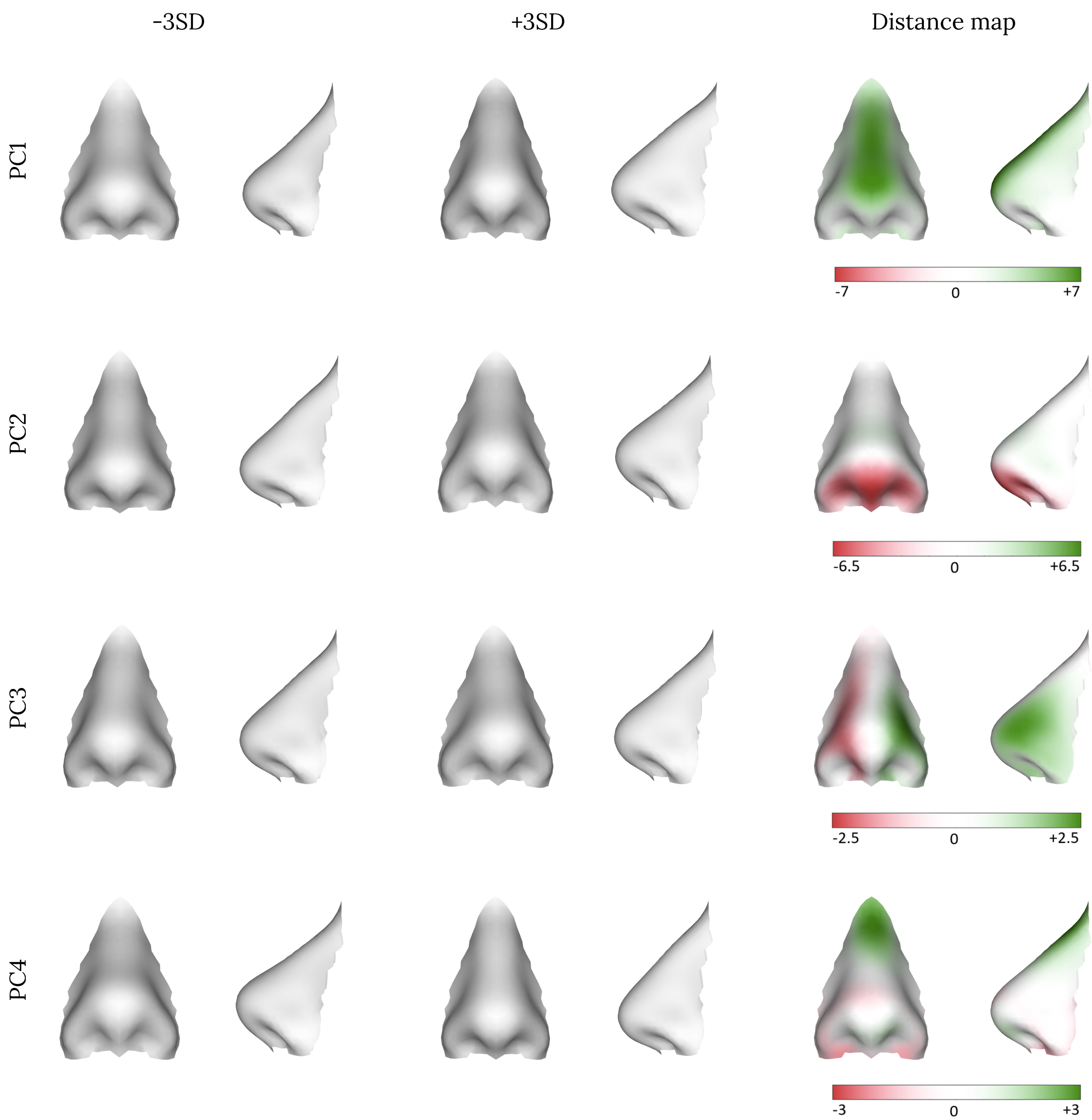

FIGURE 21. Visualization of the shape variability of the nose captured in the first four principal components. The first and second column show the frontal and profile views of the reconstructed nose regions that differ three standard deviations from the mean shape in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed regions. The distances are measured in mm the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

D.2 Descriptive statistics

TABLE 26. Descriptive statistics for the cisgender male, cisgender female and transgender female group for the principal component scores of the first four principal components.

PC	Group	N	Mean	SD
PC1	Cisgender male	518	19.942	43.871
	Cisgender female	718	-14.387	44.565
	Transgender female	203	17.754	51.373
PC2	Cisgender male	518	5.000	43.806
	Cisgender female	718	-3.607	39.614
	Transgender female	203	.447	39.857
PC3	Cisgender male	518	-1.304	18.173
	Cisgender female	718	.941	16.526
	Transgender female	203	-.802	17.723
PC4	Cisgender male	518	5.383	16.674
	Cisgender female	718	-3.884	14.819
	Transgender female	203	2.516	15.114

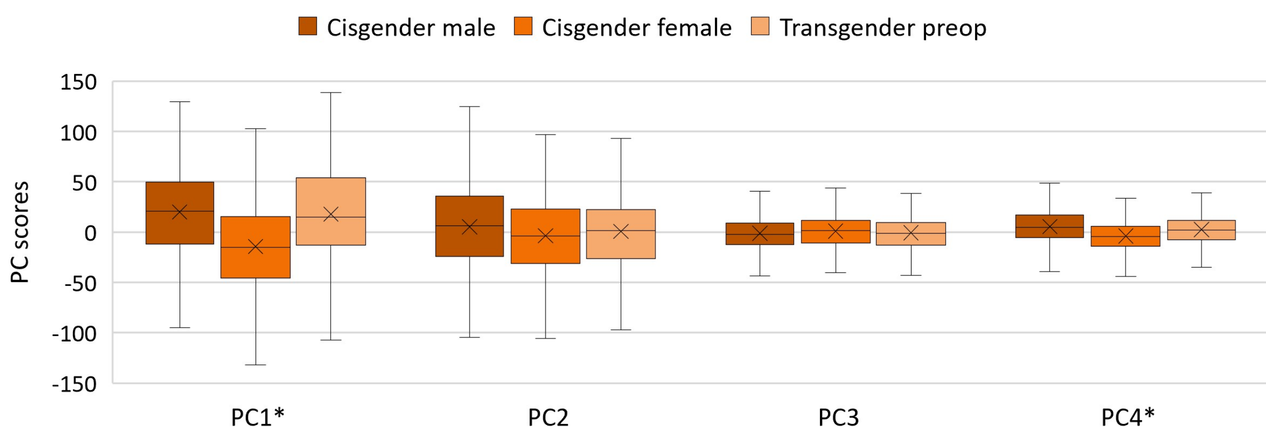

FIGURE 22. Distribution of the principal component scores for the cisgender male, cisgender female and transgender female nose region for the first 4 principal components. *The mean differences among groups is significant at the .05 level, with an effect size larger then 0.20.

E. RESULTS OF THE PCA FOR THE CHIN REGION

E.1 Visualizations of the shape variability of the chin region

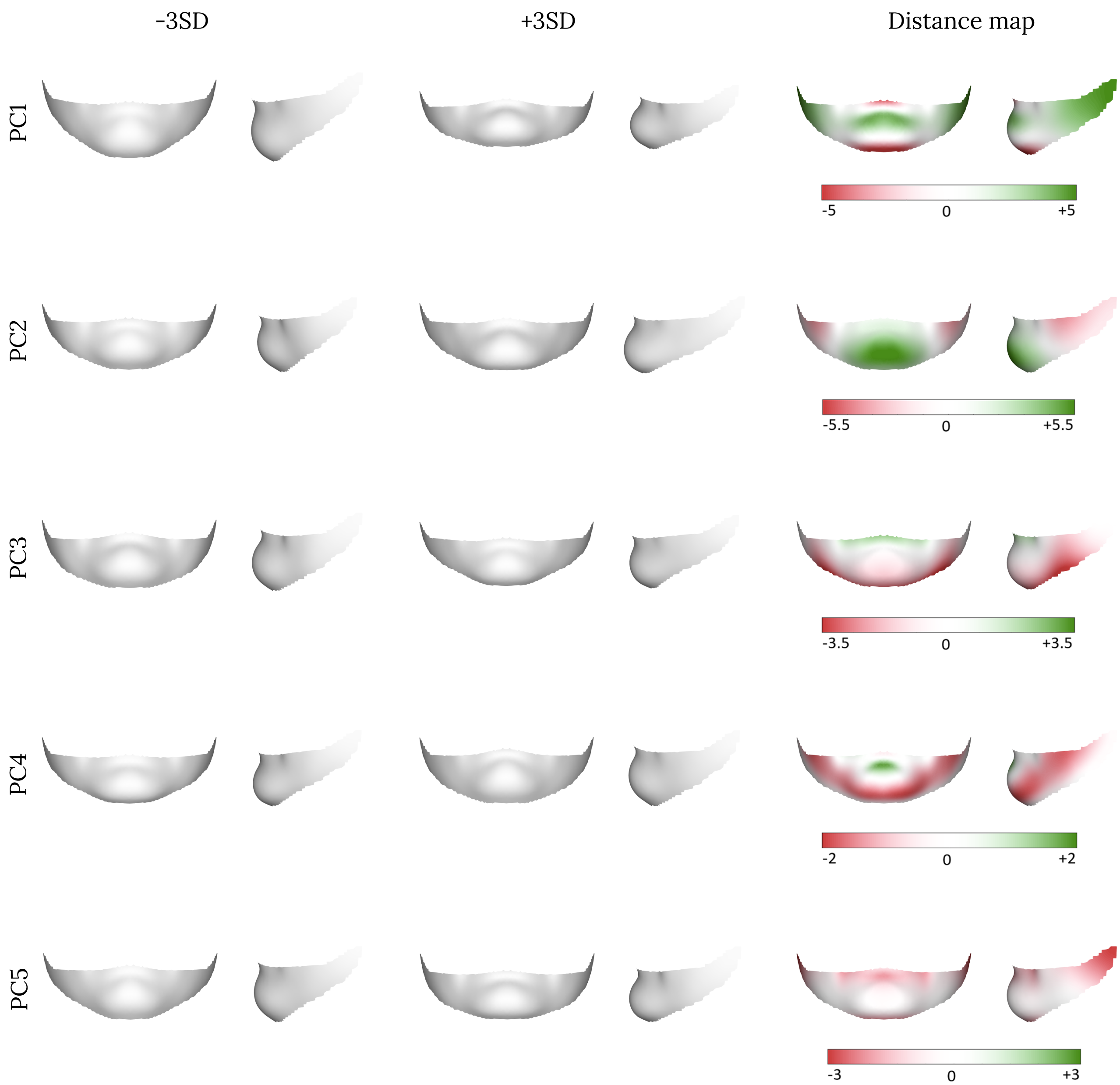

FIGURE 23. Visualization of the shape variability of the chin captured in the first five principal components. The first and second column show the frontal and profile views of the reconstructed chin regions that differ three standard deviations from the mean shape in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed regions. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

E.2 Descriptive statistics

TABLE 27. Descriptive statistics for the cisgender male, cisgender female and transgender female group for the principal component scores of the first five principal components.

PC	Group	N	Mean	SD
PC1	Cisgender male	518	-2.028	82.229
	Cisgender female	718	1.463	74.327
	Transgender female	203	9.518	85.500
PC2	Cisgender male	518	23.776	50.882
	Cisgender female	718	-17.153	49.424
	Transgender female	203	10.368	50.164
PC3	Cisgender male	518	3.432	30.473
	Cisgender female	718	-2.476	29.402
	Transgender female	203	1.754	28.585
PC4	Cisgender male	518	-1.117	30.254
	Cisgender female	718	.806	26.373
	Transgender female	203	21.654	28.308
PC5	Cisgender male	518	-7.505	22.136
	Cisgender female	718	5.414	21.026
	Transgender female	203	-7.945	22.964

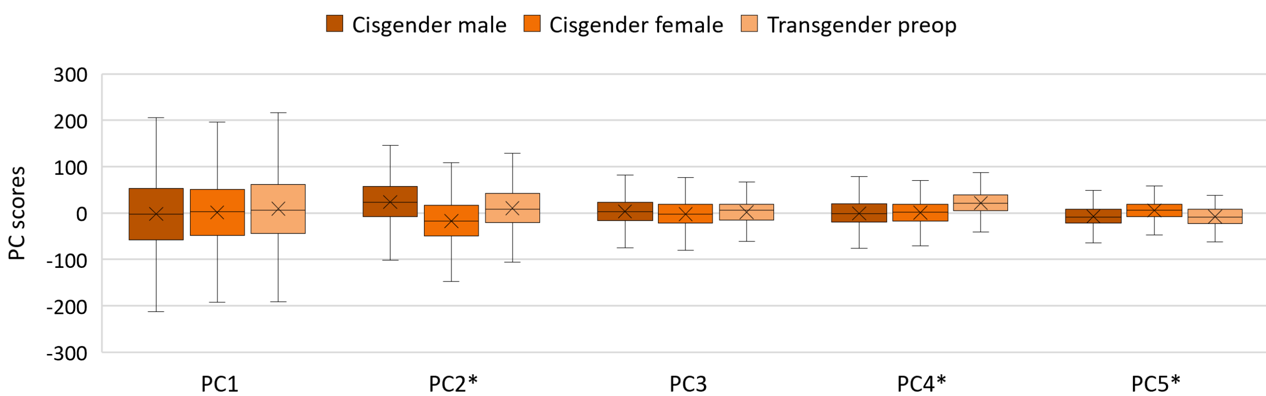

FIGURE 24. Distribution of the principal component scores for the cisgender male, cisgender female and transgender female chin region for the first 5 principal components. *The mean differences among groups is significant at the .05 level, with an effect size larger then 0.20.

F. RESULTS OF THE PCA FOR THE ZYGOMA REGION

F.1 Visualizations of the shape variability of the zygoma region

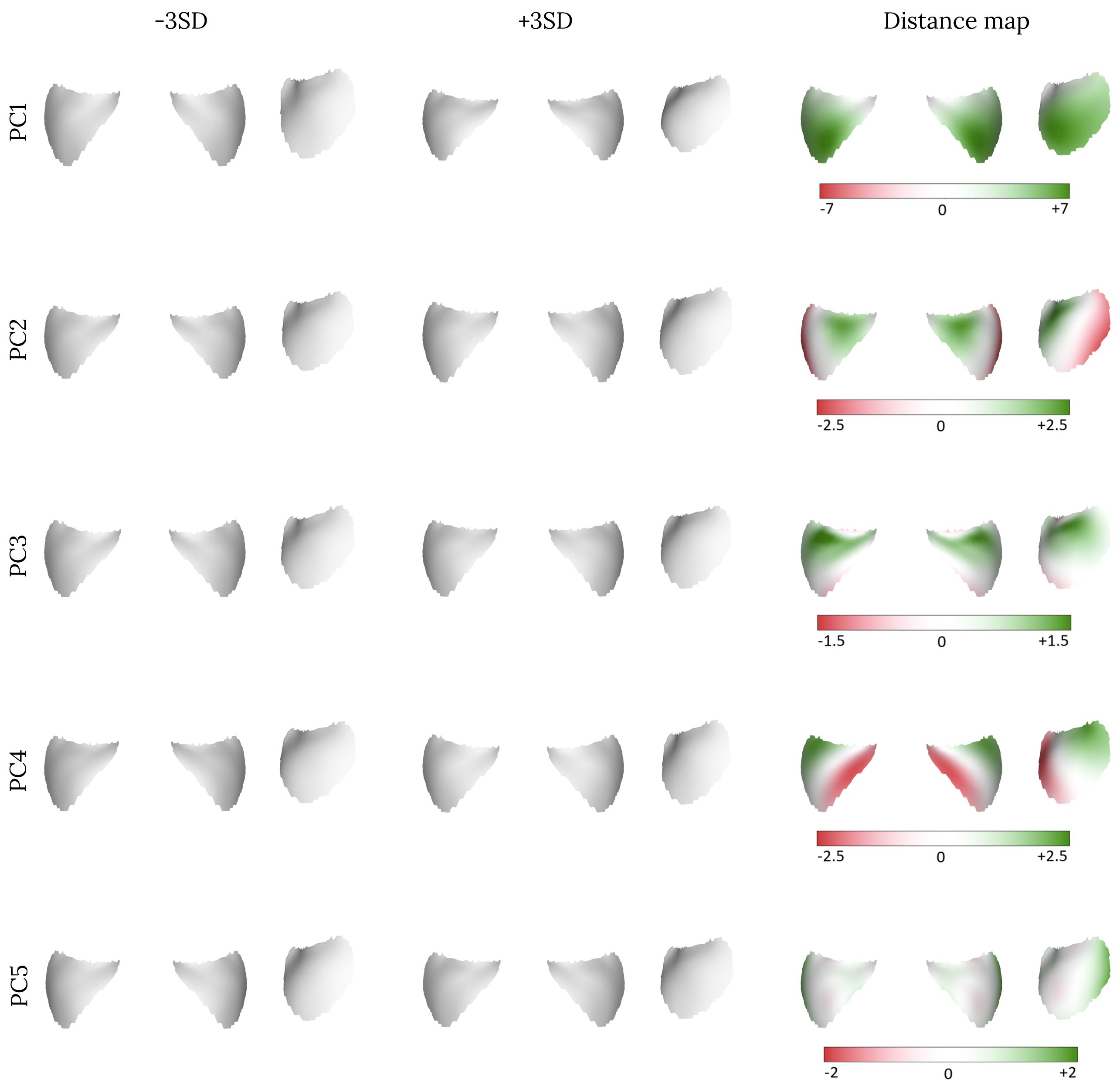

FIGURE 25. Visualization of the shape variability of the zygoma captured in the first five principal components. The first and second column show the frontal and profile views of the reconstructed zygoma regions that differ three standard deviations from the mean shape in the negative (-3SD) and positive (+3SD) direction. The third column shows the distance maps illustrating the differences between the -3SD and +3SD reconstructed regions. The distances are measured in mm in the direction perpendicular to the facial surface. (Red = pointing inwards, green = pointing outwards).

F.2 Descriptive statistics

TABLE 28. Descriptive statistics for the cisgender male, cisgender female and transgender female group for the principal component scores of the first five principal components.

PC	Group	N	Mean	SD
PC1	Cisgender male	518	-42.480	53.538
	Cisgender female	718	30.647	49.122
	Transgender female	203	-10.730	50.291
PC2	Cisgender male	518	-13.217	43.216
	Cisgender female	718	9.535	42.744
	Transgender female	203	-13.883	46.501
PC3	Cisgender male	518	6.561	29.051
	Cisgender female	718	-4.734	27.102
	Transgender female	203	20.520	26.551
PC4	Cisgender male	518	-1.897	20.433
	Cisgender female	718	1.368	19.934
	Transgender female	203	-3.776	20.207
PC5	Cisgender male	518	.814	17.274
	Cisgender female	718	-.587	17.094
	Transgender female	203	8.992	16.548

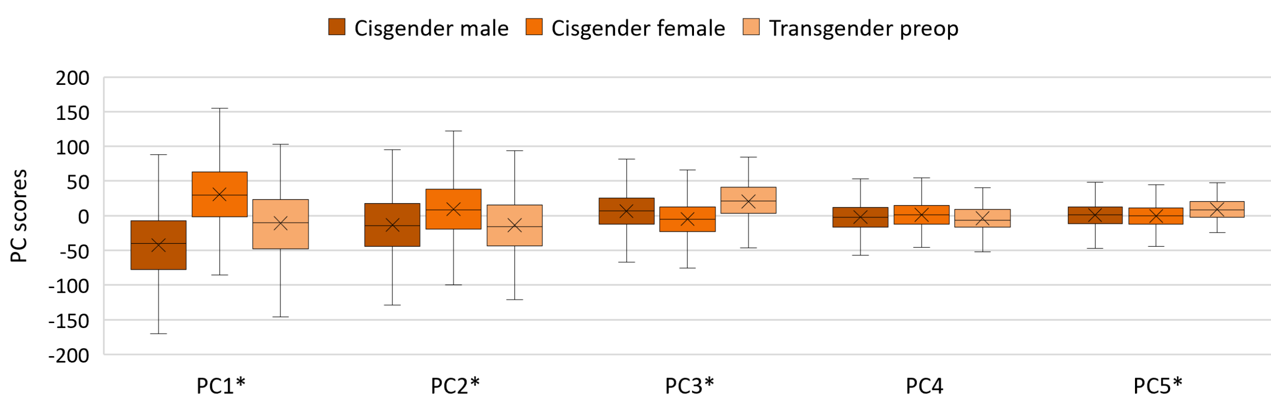

FIGURE 26. Distribution of the principal component scores for the cisgender male, cisgender female and transgender female zygoma region for the first 5 principal components. *The mean differences among groups is significant at the .05 level, with an effect size larger then 0.20.

G. RESULTS OF THE 5-FOLD CROSS-VALIDATION

TABLE 29. Results of the 5-fold cross-validation of the classification network for the full face.

	Test accuracy	AUC	Brier score	F1
Fold 1	95.56%	0.978	0.044	0.962
Fold 2	93.12%	0.980	0.069	0.941
Fold 3	91.50%	0.966	0.085	0.927
Fold 4	90.68%	0.978	0.093	0.919
Fold 5	93.93%	0.969	0.061	0.948
Mean	92.96%	0.974	0.070	0.939

TABLE 30. Results of the 5-fold cross-validation of the classification network for the forehead region.

	Test accuracy	AUC	Brier score	F1
Fold 1	88.31%	0.933	0.117	0.900
Fold 2	86.23%	0.930	0.138	0.888
Fold 3	85.83%	0.917	0.142	0.883
Fold 4	87.85%	0.932	0.121	0.897
Fold 5	86.64%	0.920	0.134	0.889
Mean	86.97%	0.926	0.130	0.891

TABLE 31. Results of the 5-fold cross-validation of the classification network for the nose region.

	Test accuracy	AUC	Brier score	F1
Fold 1	82.26%	0.899	0.177	0.855
Fold 2	79.35%	0.862	0.206	0.829
Fold 3	78.14%	0.901	0.219	0.821
Fold 4	79.76%	0.889	0.202	0.824
Fold 5	81.38%	0.898	0.186	0.839
Mean	80.18%	0.890	0.198	0.834

TABLE 32. Results of the 5-fold cross-validation of the classification network for the chin region.

	Test accuracy	AUC	Brier score	F1
Fold 1	87.90%	0.953	0.121	0.894
Fold 2	90.68%	0.954	0.093	0.922
Fold 3	89.47%	0.948	0.105	0.909
Fold 4	88.66%	0.955	0.113	0.901
Fold 5	87.04%	0.916	0.130	0.890
Mean	88.75%	0.945	0.112	0.903

TABLE 33. Results of the 5-fold cross-validation of the classification network for the zygoma region.

	Test accuracy	AUC	Brier score	F1
Fold 1	87.50%	0.955	0.125	0.896
Fold 2	88.66%	0.938	0.113	0.903
Fold 3	86.64%	0.949	0.134	0.887
Fold 4	90.28%	0.958	0.097	0.917
Fold 5	89.88%	0.943	0.101	0.913
Mean	88.59%	0.949	0.114	0.903