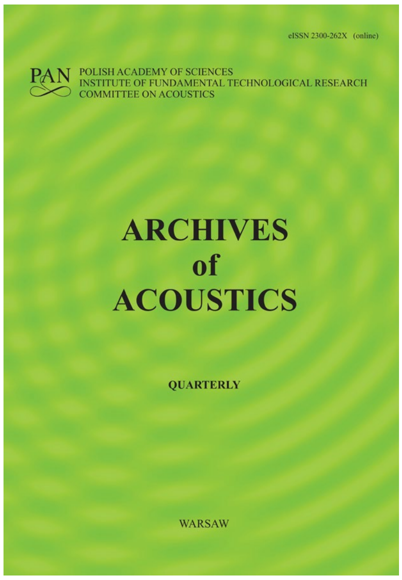
JOURNAL PRE-PROOF

This is an early version of the article, published prior to copyediting, typesetting, and editorial correction. The manuscript has been accepted for publication and is now available online to ensure early dissemination, author visibility, and citation tracking prior to the formal issue publication.

It has not undergone final language verification, formatting, or technical editing by the journal's editorial team. Content is subject to change in the final Version of Record.

To differentiate this version, it is marked as "PRE-PROOF PUBLICATION" and should be cited with the provided DOI. A visible watermark on each page indicates its preliminary status.

The final version will appear in a regular issue of *Archives of Acoustics*, with final metadata, layout, and pagination.



Title: Dimensions of the Face, Head, and Neck Affect Acoustic Parameters in Polish Speakers

Author(s): Łukasz Pawelec, Kamila Słowik, Anna Lipowicz

DOI: https://doi.org/10.24423/archacoust.2026.4173

Journal: *Archives of Acoustics*

ISSN: 0137-5075, e-ISSN: 2300-262X

Publication status: In press

Received: 2025-01-07

Revised: 2025-10-04

Accepted: 2026-01-02

Published pre-proof: 2026-01-29

Please cite this article as:

Pawelec Ł., Słowik K., Lipowicz A. (2026), Dimensions of the Face, Head, and Neck Affect Acoustic Parameters in Polish Speakers, *Archives of Acoustics*, https://doi.org/10.24423/archacoust.2026.4173

Dimensions of the Face, Head, and Neck Affect Acoustic Parameters in Polish Speakers

Łukasz Pawelec https://orcid.org/0000-0001-9406-9997, Kamila Słowik, Anna Lipowicz https://orcid.org/0000-0002-9182-6953,*

Department of Anthropology, Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland

*Corresponding Author: anna.lipowicz@upwr.edu.pl

The relationships between human voice parameters and body dimensions have been previously described, but the connections between voice and face geometry remain poorly researched. This study aimed to determine the relationships between face dimensions and acoustic parameters in both sexes and examined 111 adult participants (30 males). Each participant underwent voice recording, which included five sustained vowels, along with anthropometric measurements of the neck, head, and face regions. Comparison of the voice and the head, face, and neck regions employed Pearson's correlation coefficients (r) and a multiple linear regression model. The results revealed significant relationships between head, neck, face dimensions and acoustic parameters in both sexes. Males with higher noses, greater head circumferences and wider faces tended to have lower formants and more stable voices. Females with higher head circumferences had lower formant values, and those with more substantial neck circumferences tended to have more stable voices. Also, females with increased nose height had a lower fourth formant. Moreover, females with wider faces, noses, and jaws tended to have less rough voices (lower jitter) and longer MPT. These findings may be useful for scientists and law enforcement authorities for creating algorithms that build face models based on voice signals.

Keywords: biometry; formants; fundamental frequency; pitch; personal identification.

1. Introduction

There are known associations between vocal acoustics and body dimensions (Brueckert et al., 2006; Evans et al., 2006; González, 2004, 2007; Graddol & Swann, 1983; Pawelec et al., 2022; Pisanski et al., 2014, 2016; Rendall et al., 2005), and body composition (Hamdan et al., 2012, 2013). Such relationships depend on the correlation between vocal tract length (VTL), shape, and body size (Fitch & Giedd, 1999) and the relationship between vocal tract morphology and larynx/vocal fold size and acoustic voice parameters such as fundamental or formant frequencies (Fitch, 1997; Titze, 2011). Linear physical characteristics such as body height and weight, and the circumference of the shoulders, chest, waist, and hips, as well as proportions, i.a. the waist-to-hip ratio (WHR), are crucial for describing the appearance of an individual, but they are not as significant as the face in individual identification (Young & Bruce 2011). Moreover, when given two stimuli for personal identification (face and body), people rely more often on the face (Burton et al., 1999; O'Toole et al., 2010; Robbins & Coltheart, 2012). Additionally, judges rely more on facial features, especially nose and face shape, than body build characteristics (Rice et al., 2013).

The number of published papers examining the connections between facial morphology and voice features (Bommarito et al., 2019; Macari et al., 2015, 2017; Reinheimer et al., 2021) is limited, as are the facial measurements presented in them. Based on results of those studies weak (0.2-0.3) to moderate (0.4-0.6) correlations between facial dimensions such as i.e. jaw width (go-go), face width (*zy-zy*), maxilla width (*j-j*) or mandibular length (*co-gn/co-me*) were mentioned. Evidence that facial structure significantly affects voice parameters also comes from studies comparing facial measurements of voice professionals and control subjects (Brattström et al., 1991; Wyganowska-Świątkowska et al., 2013). Professional singers tend to have larger maxilla and mandible dimensions or greater lower face height. Furthermore, evidence indicates that faces can be correctly linked to voice with a probability greater than random using static (Mavica & Barenholtz, 2013) or dynamic facial images (Kamachi et al., 2003).

Some studies examined relations between head and neck circumferences and vocal characteristics but their findings are ambiguous - some of them showed the lack of such relationships (Evans et al., 2006; Rendall et al., 2005) and others revealed a reverse and weak association between neck circumference and voice pitch ($r \sim -0.3$; Pawelec et al., 2022) or stronger connections between first three formants (F1-F3) and head circumference (*r* ~ -0.6-0.7; Reinheimer et al., 2021). Bommarito et al. (2019) found that Martin's Facial Index (face height [*n-gn*]/face width [*zy-zy*]) correlated inversely with the second (F2) and the third formant (F3) in males (*r* ~ -0.22 and *r* ~ -0.25, respectively) and inversely with the first formant (F1) and positively with the third formant (F3) in females (*r* ~ -0.31 and *r* ~ 0.27, respectively). Macari et al. (2017) observed an inverse association between mandibular width and vocal pitch (F0) or habitual pitch (HP) in female participants (*r* ~ -0.35 and *r* ~ -0.39, respectively) and inverse correlations between F0 or HP and maxillary width in male participants (*r* ~ -0.57 and *r* ~ -0.66, respectively). This study also revealed negative correlations between face width and HP in males and females (*r* ~ -0.54 and *r* ~ -0.35, respectively). The reason why such relationships between head/neck anatomical structures and voice parameters may be observed is the fact that facial morphology is associated with the height of vocal tract cavities (i.e. oral, nasal) and pharyngeal airways (Kikuchi, 2008). In addition, vocal tract structures are related to voice parameters (Fitch, 1997), so one can find collinearity between head/neck dimensions and voice quality. Studies also demonstrate that machine learning methods can estimate face geometry from voice signals (Oh et al., 2019; Wu et al., 2022).

Accounting for the aforementioned scientific reports on the strength and directions of relationships between face and voice, the present study aimed to investigate the interrelationship between head, face, and neck dimensions and various acoustic parameters. The approach used two voice stimuli of five sustained vowels and a short sentence without emotional overtones, not just sustained vowel phonation as used in similar studies (Bommarito et al., 2019; Macari et al., 2015, 2017; Reinheimer et al., 2021). Moreover, anthropometric measurements determined the facial dimensions of live persons using calipers, and measurements did not rely on lateral radiographs (Macari et al., 2015, 2017; Reinheimer et al., 2021), other indicators (Bommarito et al., 2019) or photogrammetric technology (Lucas et al., 2023). According to Krauss et al. (2002) to identify criminals making anonymous threatening or blackmailing calls, law enforcement agencies consult voice and speech analysis experts to identify the characteristics of the speaker. Previous research attempts to estimate facial features based solely on voice signal (Li et al., 2023; Ning et al., 2021; Wen et al., 2021; Zheng et al., 2021) but they use some algorithms to match voice and face without showing the strongest determinants of acoustic parameters and their findings are inconsistent. Therefore, the current study focused on finding the head and facial features most strongly associated with voice parameters and determined the direction of these relationships. The relationships revealed could help future researchers to develop an algorithm to identify the dimensions of a speaker's face solely from a recording of their voice.

2. Material and methods

2.1. Participants

Study subjects included 135 participants (40 males) from Wroclaw, Lower Silesia, Poland. The study group consisted of students from Wroclaw University of Environmental and Life Sciences, Faculty of Biology and Animal Science, and adult inhabitants of Wroclaw invited to the research. All volunteers were examined at the same time of day (9-12 AM) and under the same conditions (the same silent room, angle, and distance from the recorder, with the same sound recording equipment used). All participants filled in a preliminary questionnaire containing inclusion/exclusion criteria, basic questions (sex, date of birth), questions on all the possible factors that could affect their acoustic parameters (history of trauma and surgery of the head and neck regions, speech defects, hearing deficits, and occlusion defects), any illness during the examination, use of cigarettes or e-cigarettes, significant alcohol consumption on the day prior, use of hormonal drugs such as anabolic steroids, use of growth hormones or hormonal contraception, and history of voice work as a teacher, sales representative, professional, or amateur singer. Females also answered questions about their current menstrual cycle, pregnancy, and menopause. No participants declared a history of head and neck trauma or surgery, hearing or speech defects, or voice work. However, excessive smoking excluded eight males and six females, while illness on the examination day excluded two males and one female. Due to the influence of sex hormones on the receptors of vocal fold epithelium (Newman et al., 2000; Voelter et al., 2008) and the consequent impact on human voice parameters (Abitbol et al., 1999; Dabbs & Mallinger, 1999; Evans et al., 2008; Raj et al., 2010), the use of hormonal contraceptives (n=4) and being in the fertile phase of the menstrual cycle (n=3) excluded seven females from further examination. This left a total of 30 males and 81 females for further study.

Research was conducted in accordance with the Declaration of Helsinki. The personal information of all participants was anonymized by giving each of them an individual anonymous code. This study was approved by the Bioethics Committee at the Wroclaw Medical University (consent number: KB - 25/2021).

2.2. Voice recording and analysis procedure

The voice recording of each participant used the same standardized conditions, and the soundtrack recorded was five vowels (/ɑː/, /ɛː/, /iː/, /ɔː/, /uː/) sustained for three seconds, with a one-second break between them. This method is one of the most often used ones in such studies (i.e. see Pisanski et al., 2014) but there are also other variants for example bVt context words (bat, bet, beet, bot, boot) or whole sentences. Speakers were asked to announce the vowels using a comfortable pitch and loudness. The equipment used included a Shure SM 58 SE dynamic cardioid microphone (bandwidth 50 Hz - 15 kHz) connected to an IMG Stageline MPA-202 amplifier and the soundcard from a Dell Latitude E6400 computer. Each participant recorded while in front of a microphone on a height-adjusted tripod, with the distance between the tip of the mouth and the recording device set at 15 cm and an angle of 0°. Standardization of the measurement conditions was achieved using a Mozos Mshield acoustic cabin. The sampling frequency was equal for all recordings and amounted to 44.1 kHz (16-bit resolution), and all tracks were saved as uncompressed format (.wav) mono sounds. The Benetech GM1351 (Benetech Poland) phonometer indicated that the value of the acoustic background of the recording room used was ~38 dB.

All recorded soundtracks were subsequently analyzed using Praat software v 3.9.2. (Boersma & Weenink, 2019). In the first instance, each vowel was analyzed separately. The middle (the most stable) part of each vowel of equal length (0.2 s) was extracted for analysis, and the "voice report" function was used to compute basic acoustic parameters such as fundamental frequency (F0 [Hz]), the lowest and strongest harmonic produced as vocal fold vibration and perceived as vocal pitch (mean F0, median F0, standard deviation [SD]-F0, min F0, and max F0), jitter [%], the degree of variation in the frequency of sound wave from period to period (local, rap, ppq5, ddp), shimme*r* [%], the degree of the amplitude of the acoustic wave variation from period to period (local, apq3, apq5, apq11, dda), mean harmonics-to-noise ratio (HNR [dB]), and the indicator of the relation of harmonics to noises in the voice (Teixeira et al., 2013; Titze, 1994). Other measurements and calculations included formant frequencies (formants; F1-F4 [Hz]), which are formed by filtering F0 in the supralaryngeal vocal tract (Fant, 1960), and their derivatives (Fn [Hz]), formant position (P_f [Z]), formant spacing (ΔF [Hz]), formant dispersion (D_f [Hz]), apparent VTL [cm] estimated from formants values, and voice intensity (loudness [dB]; see the appendix of Pisanski et al., 2014). The final values of acoustic parameters were calculated as the mean of all five vowels. Additionally, the mean maximum phonation time (MPT [s]) was measured as the maximum time of sustaining vowel /ɑː/ over three trials, with a five-second break between trials. The pitch floor was set to 75 Hz for males and 100 Hz for females, and the pitch ceiling was set to 300 Hz and 500 Hz for male and female speakers, respectively. The formant ceiling values were 5000 Hz for males and 5500 Hz for females. All acoustics parameters were computed using Praat default algorithms.

2.3. Anthropometric measurements

All measurements followed standard anthropometric procedures (Martin, 1914) with each participant subjected to two series of anthropometric measurements of the body, head, and neck. The concordance of measurements derived from two series was more than 88% ($r = 0.94$, $p < 0.001$). The body measurements included body height (cm) measured with an anthropometer to an accuracy of 0.1 cm, and body weight (kg), recorded using an electronic InBody 270 body composition analyzer device (InBody Poland) to the nearest 0.1 kg. Body mass index (BMI [kg/m^2]) was then calculated based on body height and weight values. Head and neck measurements included circumference (cm) and width-length face measurements (mm). Neck and head circumferences measurements used an anthropometric measuring tape ranging from 0 to 150 cm and a precision of 0.5 cm. Neck circumference was recorded at the laryngeal prominence ('Adam's apple'), and head circumference was measured using two points, *glabella* (*g*), the most forward point on the lower part of the forehead between the superciliary arches, and *opistocranion* (*op*), the most posterior and inferior points on the occipital bone. The face measurements were taken using a linear caliper and spreading caliper with a precision of 0.1 mm and included face height (*nasion-gnathion* [*n-gn*]), nose height (*nasion-subnasale* [*n-sn*]), nose width (*alare-alare* [*al-al*]), upper lip height (*subnasale-stomion* [*sn-sto*]), lower lip height (*stomion-supramentale* [*sto-sm*]), total lip vermillion height (*labiale superior-labiale inferior* [*ls-li*]), labial fissure length (*cheilion-cheilion* [*ch-ch*]), face width (bizygomatic diameter; *zygion-zygion* [*zy-zy*]), and jaw width (*gonion-gonion* [*go-go*]; Fig. 1). Based on measurements analyzed by Bommarito et al. (2019), Macari et al. (2015, 2017) and Reinheimer et al. (2021) and the knowledge about relation between head/face/neck dimensions and vocal tract parameters (Fitch & Giedd, 1999) above-mentioned measurements were chosen to the analyses. The rationale of choosing these head/neck measurements was i) the strongest association between them and vocal tract dimensions (i.e. oral cavity); ii) use of these facial measurements by other authors addressing similar research topics.

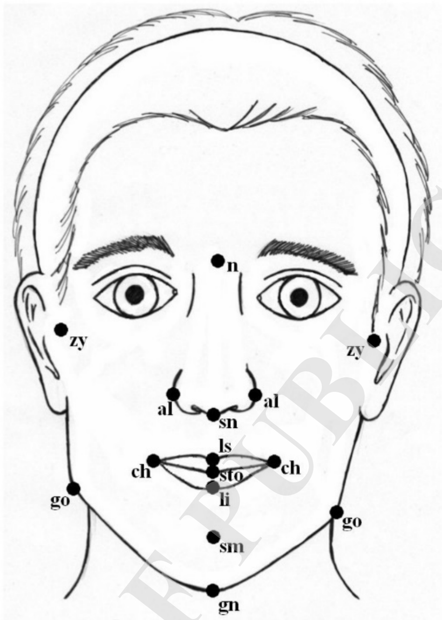


Fig. 1. Facial anthropometric landmarks used for measurements.
Source: Graja K., Król K. (2022). Antropometria i antroposkopia. Skrypt do ćwiczeń (unbundled materials of Division of Anthropology, Wroclaw University of Environmental and Life Sciences) - in self modification.

The measurement values from the two series were averaged, and the means were used for all further analyses. For each of above-mentioned head/neck measurements the intra-evaluator technical error of measurement (TEM) was calculated with reference to the method presented in the study of Perini et al. (2005). Both absolute and relative (%) error were calculated based on the equations presented below:

$$Absolute\ TEM = \sqrt{\frac{\sum d_i^2}{2n}} \qquad \text{(Eq. 1)}$$

Σd_i^2 = the sum of squared differences between 1st and 2nd measurement

n = the number of measured participants

i = the number of measurements

$$Relative\ TEM = \frac{Absolute\ TEM}{VAV} * 100\% \qquad \text{(Eq. 2)}$$

VAV – variable average value (arithmetic mean calculated on the basis of average values obtained from two measurements)

2.4. Statistical methods

All analyses employed Statistica 13.5 software (1984-2017 TIBCO Software Inc, Palo Alto, California, USA). Shapiro-Wilk *W*-test was applied to check the normality of distribution of body/facial features and acoustics parameters in male and female groups. To compare males and females, if the distribution of each variable was normal (gaussian) in both groups, independent samples *t*-test was applied and as measures of central tendency and variability mean ± standard deviation (SD) were shown. When distribution was not gaussian at least in one studied group Mann-Whitney *U*-test was applied and median and lower/upper quartiles [Q1; Q3] were demonstrated. In both cases it was shown also range (min-max). Investigating the relationship between neck and head measurements and voice parameters used Pearson's partial linear correlation due to the continuous character of the compared variables, taking age and body height as confounders because of known relations between these factors and voice parameters and head/face morphology (i.e. Fitch & Giedd, 1999; Pisanski et al., 2014, 2016; Jandová & Urbanová, 2016; Rojas et al., 2020; Pawelec et al., 2022; Cazacu et al., 2025). Furthermore, multiple linear regression models verified which head, face, and neck measurements had the most impact on the voice parameters of males and females, separately.

Several models were created for both sexes, in which selected voice parameters acted as the dependent variables (one model used one voice parameter as a dependent variable). The acoustic parameters included F0 (mean pitch), jitter parameters, shimmer parameters, HNR, intensity, P_f, ΔF, D_f, apparent VTL, and MPT. For males, explanatory variables (predictors) included head circumference and nose length (*n-sn*), while three predictors for females included lip vermillion height (*ls-li*), mandible breadth (*go-go*), face height (*n-gn*). The independent variables used for regression models were based on the highest correlation coefficients (r) for voice parameters for males and females. The collinearity of the variables used was tested by the correlation coefficients, and no redundant variables were placed into models. The highest collinearity between predictors that we allowed were at most 0.3. In each model age and body height were included as a confounders. For each model adjusted coefficient of determination (R-squared) was also computed. Only models with significant predictors were shown. Results at a significance level of $p < 0.05$ were considered statistically significant.

3. Results

3.1. Descriptive statistics

The mean participant age was 29.4 ± 12.7 years (range: 18-65 years), with males aged 5 years older than females. Moreover, males had greater values of body height, body mass, neck circumference, head circumference, face circumference, and head measurements, except for the total lip vermillion height (*ls-li*; Table 1). The intra-evaluator technical errors of measurement (absolute [cm or mm]; relative [%]) for head and face dimensions were as followed: head circumference (0.3 cm; 0.5 %), neck circumference (0.6 cm; 1.9 %), *n-gn* (2.1 mm; 1.8 %), *n-sn* (1.8 mm; 3.6 %), *al-al* (1.0 mm; 3.2 %), *sn-sto* (0.9 mm; 4.0 %), *sto-sm* (1.4 mm; 8.8 %) , *ls-li* (1.0 mm; 6.3 %), *ch-ch* (0.9 mm; 1.9 %), *zy-zy* (2.0 mm; 1.5 %), *go-go* (0.7 mm; 0.7 %).

According to the paper of Perini et al. (2005), obtained relative TEM are lower than 10 % and are acceptable.

Table 1. Physical characteristics of study participants.

Feature	Males (n=30)*	range	Females (n=81)*	range	$p^{\dagger}$
General data					
Age [y.]	27.0 [22.0; 36.0]	20 - 65	22.0 [22.0; 25.5]	18 - 64	**0.046**
Body height [cm]	180.5 ± 6.5	168.1 - 197	165.6 ± 6.5	146 - 187	**<0.001**
Body mass [kg]	89.2 [68.9; 107.9]	51.5 - 164.7	63.6 [55.5; 75.4]	47.2 - 145.2	**<0.001**
BMI [kg/m^2]	26.8 [22.2; 32.4]	17 - 46.4	23.4 [20.5; 27.2]	17.3 - 51.4	**0.037**
Head/neck measurements					
--circumferences [cm]--					
Head	57.5 ± 1.5	54 - 60	55.2 ± 1.5	52 – 58.8	**<0.001**
Neck	38.8 [36.1; 41.5]	32 - 48	32.0 [31.0; 34.0]	28 - 41	**<0.001**
--face measurements [mm]--					
n-gn	114.0 [110; 122]	65 - 131	108.5 [104.0; 112.5]	60 - 124	**<0.001**
n-sn	52.0 [54.0; 55.0]	44 - 65	49.0 [47.0; 52.5]	38 - 59	**0.005**
al-al	35.5 [33.0; 38.0]	30 - 42	32.0 [30.5; 34.0]	27 - 48	**<0.001**
sn-sto	19.5 [18.0; 22.0]	12 - 27	20.0 [18.0; 21.0]	15 - 25	0.443

sto-sm	16.0 [15.0; 19.0]	11 - 47	16.0 [14.0; 17.0]	11 - 35	0.147
ls-li	13.3 ± 4.1	5 - 23	15.1 ± 2.9	8 - 21	0.075
ch-ch	52.0 [49.0; 56.0]	42 - 62	50.0 [48.0; 52.0]	31 - 62	**0.025**
zy-zy	140.0 [132.0; 146.0]	120 -153	132.0 [128.0; 136.5]	114 -184	**0.002**
go-go	112.0 [106.0; 118.0]	98 - 125	101.0 [94.5; 105.0]	87 - 130	**<0.001**

*Mean ± SD for *t*-Student test or Median [Q1; Q3] for Mann-Whitney test, † significance of male-female difference

Table 2. Voice parameters computed from vowels for males and females.

Acoustic parameter	**Males (n=30)***	**range**	**Females (n=81)***	**range**	***p*†**
Median pitch [Hz]	110.0 [103.9; 118.3]	87.1 - 174.5	207.8 [187.7; 220.8]	158.6 - 253.4	**<0.001**
Mean pitch (F0) [Hz]	110.0 [103.9; 120.6]	87.0 - 174.6	207.8 [187.7; 221.0]	159.1 - 256.3	**<0.001**
Standard deviation [Hz]	0.9 [0.7; 2.1]	0.4 - 13.7	1.2 [1.0; 1.7]	0.3 - 14.6	**0.045**
Minimum pitch [Hz]	107.6 [102.0; 117.0]	84.5 - 172.2	205.8 [185.9; 216.7]	156.3 - 248.6	**<0.001**
Maximum pitch [Hz]	111.6 [105.6; 122.0]	87.7 - 177.0	209.5 [191.0; 224.0]	163 - 285.2	**<0.001**
Jitter (local) [%]	0.43 [0.35; 0.62]	0.23 - 2.3	0.38 [0.29; 0.46]	0.16 - 0.79	**0.013**
Jitter (rap) [%]	0.21 [0.16; 0.24]	0.12 - 1.32	0.21 [0.16; 0.26]	0.08 - 0.45	0.950
Jitter (ppq5) [%]	0.25 [0.2; 0.32]	0.15 - 1.67	0.22 [0.17; 0.25]	0.09 - 0.43	**0.006**
Jitter (ddp) [%]	0.63 [0.48; 0.73]	0.36 - 3.97	0.63 [0.48; 0.77]	0.25 - 1.36	0.886

Shimmer (local) [%]	3.17 [1.94; 4.93]	0.83 - 8.33	2.41 [1.83; 3.59]	0.77 - 7.22	0.086
Shimmer (apq3) [%]	1.55 [1.01; 2.2]	0.40 - 3.97	1.3 [0.96; 1.82]	0.40 - 3.42	0.148
Shimmer (apq5) [%]	1.94 [1.23; 2.71]	0.55 - 6.41	1.4 [1.06; 2.08]	0.45 - 4.60	**0.041**
Shimmer (apq11) [%]	2.98 [1.83; 3.95]	0.83 - 5.52	1.92 [1.45; 2.71]	0.57 - 6.74	**0.015**
Shimmer (dda) [%]	4.66 [3.04; 6.59]	1.21 - 11.92	3.9 [2.88; 5.47]	1.21 - 10.25	0.149
Mean harmonics-to-noise ratio [dB]	18.1 ± 3.3	11.5 - 23.9	22.4 ± 2.8	15.8 - 28.8	**<0.001**
Intensity [dB]	82.6 [75.2; 87.5]	63.9 - 90.8	78.9 [72.7; 84.7]	63.0 - 89.7	0.102
F1 [Hz]	570.3 ± 74.5	463.4 - 801.9	590.3 ± 60.8	404.2 - 718.7	0.152
F2 [Hz]	2771.2 [2685.2; 2891.3]	1268.8 - 2190.2	1559.5 [1519.1; 1648.3]	1306.4 - 1981.0	0.290
F3 [Hz]	2772.5 ± 162.3	2393.9 - 3079.8	2901.0 ± 144.5	2561.1 - 3194.6	**<0.001**
F4 [Hz]	3730.3 ± 163.5	3425.6 - 4072.7	4012.0 ± 196.2	3561.7 - 4461.6	**<0.001**
F*n* [Hz]	2160.8 [2090.1; 2248.2]	2008.9 - 2422.5	2276.3 [2209.8; 2336.1]	2089.1 - 2462.8	**<0.001**
P_f [Z]	-0.8 [-1.5; 0.0]	-2.2 - 1.6	0.3 [-0.4; 0.8]	-1.5 - 2.0	**<0.001**
ΔF [Hz]	1101.9 ± 80.2	965.8 - 1270.9	1137.3 ± 52.6	1018.7 - 1273.6	**0.030**
D_f [Hz]	1053.3 ± 50.7	962.0 - 1154.4	1140.6 ± 68.4	986.5 - 1324.4	**<0.001**
apparent VTL [cm]	15.3 ± 1.1	13.3 - 17.5	14.8 ± 0.7	13.2 - 16.6	**0.019**
MPT [s]	20.0 [12.1; 26.8]	4.0 - 43.8	10.6 [7.1; 14.3]	2.4 - 30.2	**<0.001**

*Mean ± SD for *t*-Student test or Median [Q1; Q3] for Mann-Whitney test, † significance of male-female difference.

The mean pitch (F0) computed for males was 110.0 Hz, and for females, it was 207.8 Hz. Also, formants and their derivative values and HNR values were lower for males. On the other hand, instability (jitter and shimmer), intensity, and apparent VTL had higher values for males. MPT values were approximately doubled in males compared to females (Table 2). Most head/face/neck features as well as acoustics parameters significantly differed between the sexes (Tables 1-2) therefore all subsequent analyses were conducted separately for each sex group.

3.2. Head and neck dimensions vs. acoustic parameters from vowels - partial correlations

For male participants, head circumference positively correlated with voice parameters from the shimmer parameters group, meaning that a higher head circumference is associated with higher amplitude of sound wave variation from period to period (less stable equates to a more hoarse voice) regardless of age and body height. Additionally, males with larger head circumferences had lower values of HNR (which also means a less stable voice), but this relationship did not reach statistical significance. Males with greater nose height (*n-sn*) had a lower value of the second formant (F2, $r = -0.43$, $p = 0.023$) but a higher value of apparent VTL ($r = 0.41$, $p = 0.03$; Fig. 2.). Meanwhile, males with greater face width did not have lower values of HNR ($r = -0.36$, *ns*) and higher values of the third formant (F3; $r = 0.28$, *ns*). Also, male subjects with increased nose height (*n-sn*) had shorter MPT, though this relationship was not statistically significant ($r = -0.36$, $0.05 < p < 0.1$; Table 3).

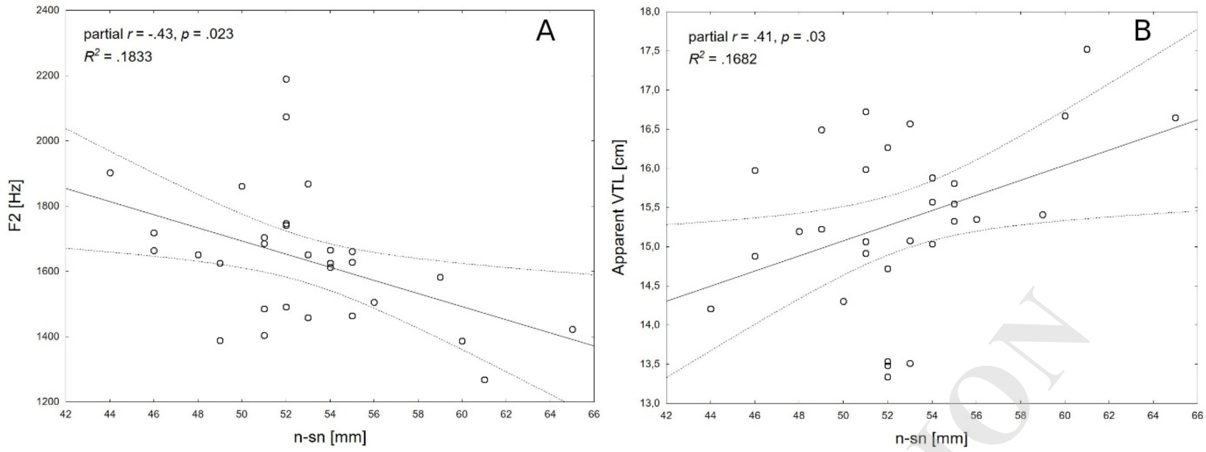


Fig. 2. Nose height (*n-sn*) and voice formants (from vowels) partial correlations controlling age and body height as confounders: a. second formant (F2): $r = -0.43$, $p = 0.023$, b. apparent VTL: $r = 0.41$, $p = 0.03$. Male participants.

For female subjects, more significant relationships between head and face dimensions and acoustic parameters extracted from vowels were found (Table 3). Female participants with greater neck circumference had lower values of shimmer parameters (apq3, dda, both $r = -0.27$, $p = 0.018$ and apq5, apq11 both $r = -0.25$, $p = 0.032$). Furthermore, females with greater face height (*n-gn*) had lower voice intensity (loudness; $r = -0.46$, $p < 0.001$). Additionally, a higher nose (*n-sn*), lower fourth formant (F4), and longer apparent VTL had a significant positive relationship with lip vermilion height (*ls-li*) and voice parameters. A higher value of this characteristic was positively correlated with shimmer parameters, which means that higher lip vermillion and a higher unstable (hoarse) voice was related to a greater value of face width (*zy-zy*), with lower jitter parameter values (rap, ppq5, ddp). Also, jaw width (*go-go*) showed similar relationships with jitter parameters, but these correlations did not reach statistical significance. Meanwhile, female participants with wider jaws (*go-go*) had higher values of MPT ($r = 0.25$, $p = 0.032$; Table 3, Fig. 3).

Most of the above correlations were weak (~0.2-0.3) as it was expected, but some of them (i.e. nose height (*n-sn*) vs. F2) showed moderate effect size for male participants.

Table 3. Head/neck circumferences and face measurements vs. voice parameters from vowels. Pearson's correlation coefficients for males (n=30) and females (n=81).

Acoustic parameter	Sex	Circumferences [cm]		Face measurements [mm]								
		Head	**Neck**	**n-gn**	**n-sn**	**al-al**	**sn-sto**	**sto-sm**	**ls-li**	**ch-ch**	**zy-zy**	**go-go**
Mean pitch (F0) [Hz]	M	-.02	-.03	.12	.15	.27	-.21	.03	-.25	.008	.12	.15
	F	.06	.08	-.04	-.01	-.10	.03	-.20	-.09	.08	.04	.08
Jitter [%]												
- local	M	.15	.12	-.05	.11	-.15	-.02	.18	-.35	.10	.21	.20
	F	-.10	-.05	-.07	-.09	.06	-.10	.09	.06	.03	-.12	-.14
- rap	M	.23	.31	.05	-.01	-.07	-.02	.08	-.16	.11	.31	**.45**
	F	-.09	-.001	-.13	-.13	.03	-.10	.11	.04	.05	-.14	-.19
- ppq5	M	.23	.11	.04	.07	-.10	.06	.10	-.20	.13	.18	.21
	F	-.10	-.08	-.08	-.04	.03	-.08	.04	.08	.03	**-.22**	-.19
- ddp	M	.23	.31	.05	-.006	-.007	-.002	.12	-.06	.01	.001	.31
	F	-.04	.04	-.12	-.11	.04	-.07	.13	.06	.10	**-.23**	-.15
Shimmer [%]												
- local	M	**.41**	.21	-.11	.08	.04	.05	.03	-.18	**.36** [b]	.20	.21
	F	-.17	**-.27**	.12	.14	-.01	.17	.01	**.29**	-.03	.007	**-.22** [b]
- apq3	M	**.43**	.26	-.02	.03	.13	.03	.06	-.21	**.38**	.26	.33

							continued					
	Sex	**Circumferences [cm]**		**Face measurements [mm]**								
		Head	**Neck**	**n-gn**	**n-sn**	**al-al**	**sn-sto**	**sto-sm**	**ls-li**	**ch-ch**	**zy-zy**	**go-go**
	F	-.18	**-.27**	.10	.07	-.03	.18	.02	**.25**	-.04	-.03	**-.22** [b]
- apq5	M	**.42**	.22	-.006	.01	.19	.01	.04	-.19	.11	.23	.24
	F	-.01	**-.25**	.11	.09	.07	.19	.03	**.29**	-.01	.03	-.20
- apq11	M	**.39**	.22	-.19	.01	.01	.09	-.02	-.21	**.38**	.24	.14
	F	-.14	**-.25**	.16	.12	-.11	.15	.01	**.32**	.002	.06	-.14
- dda	M	**.43**	.26	.02	.03	.13	.03	.06	-.21	.21	.26	.33
	F	-.18	**-.27**	.10	.07	-.03	.18	.02	**.25**	-.04	-.03	**-.22** [b]
Harmonic-to-noise ratio [dB]	M	-.32	**-.40**	.03	.08	-.19	-.07	.03	.26	-.29	-.36	-.28
	F	.12	.17	-.19	-.03	-.07	-.14	.04	-.13	.03	.13	.13
Intensity [dB]	M	-.17	-.08	-.005	-.15	.18	-.15	.01	-.21	-.29	-.12	-.16
	F	.06	.08	**-.46**	-.21	.10	-.11	.11	-.11	-.07	-.13	.03
Formants and their derivatives												
F1 [Hz]	M	.003	.11	-.24	-.25	.22	-.05	.06	-.21	-.06	-.06	-.08
	F	-.11	-.16	.02	-.05	-.07	.14	-.07	-.08	-.06	-.08	-.10
F2 [Hz]	M	.17	.25	-.14	**-.43**	.30	.13	.02	-.26	.01	.18	.10
	F	-.15	-.04	.07	-.21	-.006	-.05	-.09	-.15	-.005	-.07	.03

continued

Acoustic parameter	Sex	Circumferences [cm]		Face measurements [mm]								
		Head	**Neck**	**n-gn**	**n-sn**	**al-al**	**sn-sto**	**sto-sm**	**ls-li**	**ch-ch**	**zy-zy**	**go-go**
F3 [Hz]	M	.03	.19	-.13	-.21	.07	-.16	-.26	-.26	.12	.28	.24
	F	-.10	-.13	.03	.11	.13	-.06	.11	-.05	.10	-.12	.01
F4 [Hz]	M	.05	.17	.08	-.09	.15	.09	-.06	.03	.07	.13	.01
	F	-.003	.08	-.16	**-.24**	.11	.03	-.20	-.17	-.08	-.08	.12
P_f [Z]	M	.10	.25	-.12	-.33	.24	.03	-.09	-.27	.06	.21	.12
	F	-.11	-.05	-.05	-.15	.10	.002	-.13	-.13	-.01	-.13	.08
ΔF [Hz]	M	.09	.22	-.19	-.35	.27	.01	-.03	-.29	.01	.13	.06
	F	-.15	-.13	.003	-.14	.02	.06	-.10	-.10	.05	-.12	-.07
D_f [Hz]	M	.05	.12	.20	.03	.04	.12	.25	.14	.10	.17	.05
	F	.03	.12	-.15	-.17	.12	-.01	-.07	-.05	-.05	-.05	.15
Apparent VTL [cm]	M	-.11	-.25	.22	**.41**	-.28	-.04	.02	.27	-.02	-.16	-.08
	F	.14	.12	.01	.14	-.03	-.07	.12	.14	.04	.14	.07
MPT [s]	M	.08	-.006	-.03	-.36	.12	.001	-.35	-.08	.27	.22	.12
	F	.08	.19	.17	.08	-.04	.07	.04	-.03	-.07	-.07	**.25**

M – males, F – females; $p < 0.05$ results were **bolded,** [b] – borderline significance ($0.06 > p > 0.05$)

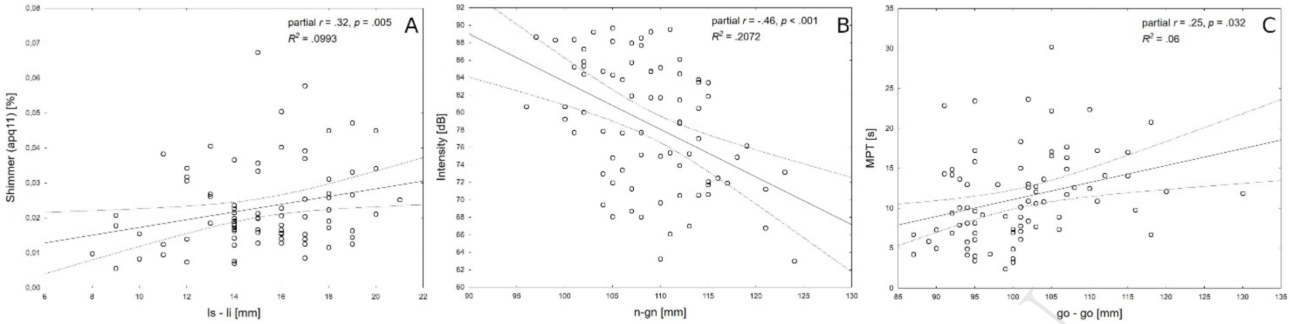


Fig. 3. a. Lips vermilion height (*ls-li*) and shimmer (apq11): $r = 0.32$, $p = 0.005$, b. face height (*n-gn*) and voice intensity (loudness): $r = -0.46$, $p < 0.001$, c. jaw width and maximum phonation time (MPT): $r = 0.25$, $p = 0.032$ partial linear correlations controlling age and body height as confounders. Female participants.

3.3. Head and neck dimensions vs. acoustic parameters for vowels: multiple regression models

For males, multiple regression modeling indicated a significant positive relationships of head circumference and shimmer (local), meaning that those with larger head circumferences had more unstable (hoarse) voices ($\beta = 0.42$, $t = 2.44$, $p = 0.022$; see Table 4, model 1). The second model revealed a significant correlation between nose height with apparent VTL, so males with higher noses had longer estimated vocal tracts ($\beta = 0.42$, $t = 2.39$, $p = 0.024$; see: Table 4, model 2).

For females, the first multiple regression model revealed that lip vermillion height significantly correlated with shimmer (apq11), which mean a larger value of *ls-li* ($\beta = 0.35$, $t = 2.99$, $p = 0.004$), was connected to a more hoarse voice (see: Table 4, model 3). The second model showed a borderline significance reverse correlation of face height (*n-gn*) and voice intensity (loudness), which means that women with higher faces had quieter voices ($\beta = -0.23$, $t = -1.96$, $p = 0.054$; see: Table 4, model 4). The last significant model created for females showed borderline significance correlation between jaw width (*go-go*) and maximum phonation time

(MPT; $\beta = 0.24$, $t = 1.98$, $p = 0.052$). This associations means that women with wider jaws had longer maximum time of phonation (see Table 4, model 5).

Table 4. Multiple general regression models (GRM) of acoustic parameters (dependent variables) and head/neck measurements (predictors) for males (n=30) and females (n=81).

Sex	Regression model	Predictors	β	SE_β	t	p
Males	*1. Dependent variable:* **Shimmer (local) [%]**	Intercept	-	-	-1.73	0.097
	$R^2_{Adj.} = 13.29\%$; $\varepsilon = 0.017$	Head circ. [cm]	0.41	0.18	2.25	**0.034**
	$F = 3.22$; $p < 0.06$	n-sn [mm]	0.07	0.19	0.38	0.704
		Age [y.]	0.09	0.19	0.46	0.650
		Body height [cm]	0.02	0.18	0.10	0.918
	2. Dependent variable: **apparent VTL [cm]**	Intercept	-	-	1.15	0.260
	$R^2_{Adj.} = 12.42\%$; $\varepsilon = 0.99$	Head circ. [cm]	-0.12	0.18	-0.65	0.522
	$F = 3.06$; $p < 0.06$	n-sn [mm]	0.42	0.18	2.28	**0.031**
		Age [y.]	-0.13	0.19	-0.70	0.489
		Body height [cm]	0.16	0.18	0.91	0.373
Females	*3. Dependent variable:* **Shimmer (apq11) [%]**	Intercept	-	-	-2.52	**0.014**
	$R^2_{Adj.} = 13.45\%$; $\varepsilon = 0.011$	ls-li [mm]	2.99	0.004	2.99	**0.004**
	$F = 3.43$; $p < 0.008$	go-go [mm]	-1.52	0.132	-1.52	0.132
		n-gn [mm]	1.84	0.070	1.84	0.070

	Age [y.]	0.33	0.13	2.44	**0.017**
	Body height [cm]	0.31	0.12	2.71	**0.008**
4. Dependent variable: **Intensity [dB]**	Intercept	-	-	4.44	**<0.001**
$R^2_{Adj.} = 0.16\%$; $\varepsilon = 7.15$	ls-li [mm]	-0.13	0.12	-1.07	0.288
$F = 1.02$; $p < 0.410$	go-go [mm]	0.06	0.13	0.48	0.632
	n-gn [mm]	-0.23	0.12	-1.96	**0.054** [b]
	Age [y.]	-0.20	0.14	-1.37	0.175
	Body height [cm]	-0.09	0.13	-0.73	0.467
5. Dependent variable: **MPT [s]**	Intercept	-	-	-2.18	**0.033**
$R^2_{Adj.} = 8.59\%$; $\varepsilon = 5.52$	ls-li [mm]	-0.02	0.12	-0.16	0.875
$F = 2.47$; $p < 0.040$	go-go [mm]	0.24	0.12	1.98	**0.052** [b]
	n-gn [mm]	0.14	0.11	1.23	0.221
	Age [y.]	0.06	0.14	0.46	0.650
	Body height [cm]	0.20	0.12	1.67	0.099

β – standardized regression coefficient; SE_{β} – standard error of β; $R^2_{Adj.}$ – adjusted R-squared of a model; ε – standardized random error of a model; t – t-statistic of β showing a significance of a predictor; F – Fisher test value showing a significance of the whole model. The results significant on the $p < 0.05$ level were **bolded,** [b] – borderline significance ($0.06 > p > 0.05$)

4. Discussion

Significant relationships between head, neck, and face dimensions and voice parameters are apparent for both sexes. The strongest relationships for males occurred between shimmer parameters and head circumference, F2, nose height and apparent VTL, as well as, between jitter (rap) and jaw width (*go-go*). As such, males with larger heads had more hoarse (unstable) voices, and those with higher noses had lower values of F2 and longer apparent VTL. Males with wider jaws had greater voice roughness. For females, the strongest connections existed for face height and voice intensity, neck circumference and lip vermilion height with shimmer parameters, nose height with F4 and jaw width with maximum phonation time (MPT). These results suggest that females with smaller face height have higher voice intensity. Also, female participants with greater neck circumferences and lower lip height had more stable (less hoarse) voices. Finally, females with longer noses had lower fourth formant value and those with wider jaws – longer MPT. Furthermore, the relationships between lip vermilion height and shimmer remained significant after applying multiple linear regression models. There were no significant relationships between F0, formant frequencies (F1-F4), derivatives (P_f or D_f), and head or neck circumferences for males and females, but these relationships were found for shimmer parameters. These results confirm partially the findings of Evans et al. (2006), who found the lack of such relationships for males, and are supported by the results of Rendall et al. (2005), who found no significant connections for both sexes. However, Pawelec et al. (2022) found significant associations between neck circumference and formants and F0 for males, even after applying multiple regression models.

For females, the strongest relationships based on multiple linear regression model were found for jaw width and MPT. Interestingly, no significant associations were found between voice pitch (F0) and head, neck, and face dimensions. Macari et al. (2017) found a significant negative correlation between face width (*zy-zy*) and habitual pitch (F0 for the sentence) in males and inverse correlations between habitual pitch and total face height (*nasion-menton*, *n-me*, measurement corresponding to our *n-gn*) and jaw width for females. Moreover, one study showed a negative relationship between Martin's Facial Index (*n-gn*/*zy-zy*) and F3 in males and a positive correlation for females (Bommarito et al., 2019). Consequently, females with higher and narrower faces and males with shorter and wider faces had higher values of F3. Wu et al. (2022) found that voice may indicate wider or thinner faces and stated that "*the best indicative attribute voice can hint is the head width*." This finding is consistent with the results of the current study, which indicate many significant correlations between acoustic parameters and face width measurements (*go-go*, *zy-zy*, *ch-ch*) in males and females.

Connections between face morphology and voice parameters found in this study are logical and result from the anatomy and physiology of the speech apparatus. It is known that the length of the vocal tract is closely related to body size and shape, including body height and mass (Fitch & Giedd, 1999) and taller and heavier individuals tend to have lower voices, expressed by lower values of fundamental and formant frequencies (Evans et al., 2006; González 2004, 2007; Pawelec et al., 2022; Pisanski et al., 2014; Rendall et al., 2005). Meanwhile, larger individuals have a bigger larynx and a longer vocal tract (Fitch & Giedd, 1999), and such anatomical structures affect the voice via longer vocal folds, lower F0, longer vocal tract, and lower formant frequencies (Fitch, 1997; Titze, 2011). Therefore, a reverse relationship between body size and voice parameters is observed (Brueckert et al., 2006; Rendall, 2005), which is also pointed out by subjective assessment of body build done by judges (Pawelec et al., 2023). A positive correlation was also found between head length (*g-op*), face height (*n-gn*), and growth indexes of the vocal tract structures during early childhood (Voperian et al., 1999). Thus, significant associations between head and face dimensions and acoustic parameters found in this study seem justified and logical. Moreover, all presented results show pure effect of head/neck/face dimensions on vocal characteristics, not affected by body size and age. Size (expressed by body height) ang metrical age of participants were controlled as confounders in partial correlations and regression models, thus their apparent impact on head/neck/face – voice relations is reduced.

Relationships between jaw width and acoustic parameters in females seem difficult to explain at first glance. However, females with greater face width tend to have lower jitter values (more stable voices). Moreover, females with wider mandibles have lower shimmer parameters (less hoarse voice) and greater MPT. It was also found that female vocalists had significantly greater face dimensions and mandible width than a control group of non-vocalists (Wyganowska-Świątkowska et al., 2013). These results may suggest a role for face and jaw morphology in voice emission (especially articulation) and speech processes. The larger jaw size is likely associated with a stronger development of the muscles attached to the mandible and greater development of the entire stomatognathic apparatus. Consequently, this affects the biomechanics of these structures during voice emission. Some support for this thesis is provided by results showing that nonsingers using a low mandible maneuver (LMM), a technique used in top-ranked professional singers to enhance vocal output by altering oral, pharyngeal, and laryngeal configurations, had an increase in voice intensity (sound-pressure level [SPL]) and lower F1 and F2 (Mercer & Lowell, 2020).

The significant negative relationships between face height, nose height, intensity, and formants, and the positive correlations between those dimensions and apparent VTL in males and females can be explained by knowledge of the speech apparatus anatomy. Facial morphology, such as face and nose height, is associated with the oral and nasal cavity height and the pharyngeal airways (Kikuchi, 2008). Total VTL is highly positively correlated with oral cavity length, pharyngeal cavity length, and palate height in both sexes (Roers et al., 2009). Furthermore, VTL is inversely related to formant frequencies and their derivatives (Fitch, 1997). As such, the aforementioned relationships are likely to result in an apparent correlation of face and nose height with the formant frequencies of the voice.

There were found no associations between lip vermilion height and formants or their derivatives in both sexes. This result is incomprehensible, as some study have shown a relationship between lip shape and F1 and F2 (Ladefoged et al., 1978) or a significant association between formants and lip rounding (Wood 1985). A significant relationship was also found between this mouth shape and the first two formants, with F1 closely related to mouth height and F2 corresponding to mouth width (Kim et al., 2002). Another study showed a high correlation between the F1 and mouth height but no correlation between F2 and the mouth height (Erber, 1979). In present study lip vermilion height was correlated positively with shimmer parameters but only in females.

4.1. Study limitations and future perspectives

The study was limited by the restricted number of head and neck measurements, and future studies should examine more head and face dimensions, such as head length (*glabella-opisthocranion* [*g-op*]), physiognomic face height (*trichion-gnation* [*tr-gn*]), forehead width (*frontotemporal-frontotemporal* [*ft-ft*]), and forehead height (*trichion-nasion* [*tr-n*]). Furthermore, it is essential to examine the impact of more vocal tract structures, such as the frontal sinuses, on voice features. Another limitation is the lack of head and neck imaging, such as x-ray, computed tomography, and magnetic resonance imaging, which would have allowed for the true dimensions of the vocal tracts of the subjects to be shown and highlight their influence on voice acoustic parameters. In addition, the unequal number of male and female participants makes cross-sex comparisons difficult. Some of the relationships presented in this study may not be authoritative in the perspective of practical usability because of the methodology applied in part of this work - isolated sustained vowels. Correlations with acoustic parameters computed from short sentence or spontaneous speech seem to be more valuable in ecological validity as these signals are more similar to natural conditions. Also the limitations of the study is the lack of controlling other factors which may influence the voice quality such as hormones (Damrose 2009; Kirgezen et al., 2017; Newman et al., 2000; Voelter et al., 2008), social context (Sorokowski et al., 2019), emotions (Klasmeyer & Sendlmeier 2000; Raine et al., 2019; Rothkrantz et al., 2004; Sondhi et al., 2015), used stimulants (Byeon & Cha 2020; Moreira et al., 2015) or various aspects of physical and mental health status (Arnocky et al., 2018).

Finally, the dimensions of the palate, including height, width, and length, could be included in future studies, as there are indications that palate shape influences voice type in opera singers (Bottalico et al., 2021; Marunick & Menaldi, 2000).

5. Conclusions

Some significant correlations were found between head and face morphology and vocal acoustic parameters in males and females, which suggests a simple relationship between vocal tract structure and function and voice emission. Indeed, both facial width and length measurements had significant negative relationships with voice parameters (Bommarito et al., 2019; Macari et al., 2015, 2017; Reinheimer et al., 2021). Also, connections were highlighted between head and neck circumferences and acoustic parameters (Pawelec et al., 2022; Reinheimer et al., 2021). VTL and shape are associated with voice parameters, especially formant frequencies (Fitch, 1997; Story et al., 2001) and body size, height, and weight (Fitch & Giedd, 1999). Moreover, head and face size and proportions are related to body size, VTL, and shape, meaning a correlation might be observed between face and head dimensions and acoustic parameters. All observed associations were quite weak in most cases and some were moderate. These relationships may be used by forensic scientists to estimate the facial morphology of offenders based on voice recordings alone (Bunker, 2017; Oh et al., 2019). It is necessary to expand this research by increasing the number of head and face measurements and sample size to allow for a meta-analysis of the relationship between facial morphology and voice acoustic parameters. Another way is to use geometric morphometrics' method taking into account not size of head/face but their shape (i.e. some indices describing face geometry). The results of such a study would enable future investigators to identify perpetrators and victims based solely on audio recordings or evaluation of voice changes in people after surgical interventions of the head and face (medical and aesthetic plastic surgery). A recent study shows some degree of success in estimating facial morphology from the voice signal (Li et al., 2023).

Acknowledgements

We would like to express our gratitude to all the participants who have given their voice as well as anthropometric measurements for this research.

Funding

This work was supported by the Wrocław University of Environmental and Life Sciences (Poland) as the Ph.D. research program „Bon doktoranta SD UPWr" no N020/0008/20 from the subsidy increased by the minister responsible for higher education and science for the period 2020-2026 in the amount of 2% of the subsidy referred to Art. 387 (3) of the Act of 20 July 2018 – Law on Higher Education and Science, obtained in 2019.

Conflict of interest

None

References

1. Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. Journal of voice, 13(3), 424-446. https://doi.org/10.1016/S0892-1997(99)80048-4
2. Arnocky, S., Hodges-Simeon, C. R., Ouellette, D., & Albert, G. (2018). Do men with more masculine voices have better immunocompetence?. *Evolution and Human Behavior*, *39*(6), 602-610. https://doi.org/10.1016/j.evolhumbehav.2018.06.003
3. Boersma P, & Weenink D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.56, Available at: http://www.praat. org/. Accessed June 20, 2019.
4. Bommarito, S. (2019). Correlation between voice, speech, body and facial types in young adults. Global Journal of Otolaryngology, 20(4). https://doi.org/10.19080/gjo.2019.20.556041
5. Bottalico, P., Marunick, M. T., Nudelman, C. J., Webster, J., & Jackson-Menaldi, M. C. (2021). Singing voice quality: The effects of maxillary dental arch and singing style. Journal of Voice, 35(3), 501-e11. https://doi.org/10.1016/j.jvoice.2019.09.015
6. Brattström, V., Odenrick, L., & Leanderson, R. (1991). Dentofacial morphology in professional opera singers. Acta Odontologica Scandinavica, 49(3), 147-151. https://doi.org/10.3109/00016359109005899
7. Bunker, D. (2017). Speech2Face: reconstructed lip syncing with generative adversarial networks. Data reflexions: Thoughts and Projects, 8.
8. Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. Psychological Science, 10(3), 243-248. https://doi.org/10.1111/1467-9280.00
9. Byeon, H., & Cha, S. (2020). Evaluating the effects of smoking on the voice and subjective voice problems using a meta-analysis approach. *Scientific Reports*, *10*(1), 4720. https://doi.org/10.1038/s41598-020-61565-3

10. Cazacu, C. J., Jula, C. R., Şapte, E., Chistol, R. O., Furnică, C., Grammatikis, E., & Bulgaru-Iliescu, A. I. (2025). Morphology of facial aging: a shape-based quantification. Romanian journal of morphology and embryology= Revue roumaine de morphologie et embryologie, 66(2), 367-373.

11. Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. Personality and Individual Differences, 27(4), 801-804. https://doi.org/10.1016/S0191-8869(98)00272-4

12. Damrose, E. J. (2009). Quantifying the impact of androgen therapy on the female larynx. Auris Nasus Larynx, 36(1), 110-112. https://doi.org/10.1016/j.anl.2008.03.002

13. Erber, N. P. (1979). Real-time synthesis of optical lip shapes from vowel sounds. The Journal of the Acoustical Society of America, 66(5), 1542-1544. https://doi.org/10.1121/1.383511

14. Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. Biological Psychology, 72(2), 160–163. https://doi.org/10.1016/j.biopsycho.2005.09.003

15. Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. Physiology & Behavior, 93(4-5), 783-788. https://doi.org/10.1016/j.physbeh.2007.11.033

16. Fant, G. (1970). Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations (No. 2). Walter de Gruyter.

17. Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. The Journal of the Acoustical Society of America, 102(2), 1213-1222. https://doi.org/10.1121/1.421048

18. Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. The Journal of the Acoustical Society of America, 106(3), 1511–1522. https://doi.org/10.1121/1.427148

19. González, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. Journal of Phonetics, 32(2), 277-287. https://doi.org/10.1016/S0095-4470(03)00049-4

20. González, J. (2007). Correlations between speakers' body size and acoustic parameters of voice. Perceptual and Motor Skills, 105(1), 215-220. https://doi.org/10.2466/pms.105.1.215

21. Graddol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. Language and Speech, 26(4), 351-366. https://doi.org/10.1177/002383098302600

22. Hamdan, A. L. H., Al Barazi, R., Khneizer, G., Turfe, Z., Sinno, S., Ashkar, J., & Tabri, D. (2013). Formant frequency in relation to body mass composition. Journal of Voice, 27(5), 567-571. https://doi.org/10.1016/j.jvoice.2012.09.005

23. Hamdan, A. L., Al-Barazi, R., Tabri, D., Saade, R., Kutkut, I., Sinno, S., & Nassar, J. (2012). Relationship between acoustic parameters and body mass analysis in young males. Journal of Voice, 26(2), 144-147. https://doi.org/10.1016/j.jvoice.2011.01.011

24. Jandová, M., & Urbanová, P. (2016). The relationship between facial morphology, body measurements and socio-economic factors. Anthropological Review, 79(2), 181-200.

25. Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the face to the voice': Matching identity across modality. Current Biology, 13(19), 1709-1714. https://doi.org/10.1016/j.cub.2003.09.005

26. Kikuchi, Y. (2008). Three-dimensional relationship between pharyngeal airway and maxillo-facial morphology. The Bulletin of Tokyo Dental College, 49(2), 65-75. https://doi.org/10.2209/tdcpublication.49.65

27. Kim, T., Kang, Y., & Ko, H. (2002). Achieving real-time lip-synch via SVM-based phoneme classification and lip shape refinement. In Proceedings. Fourth IEEE International Conference on Multimodal Interfaces (pp. 299-304). IEEE. https://doi.org/10.1109/ICMI.2002.1167010

28. Kirgezen ,T., Sunter, A. V., Yigit, O., Huq, & G. E. (2017). Sex hormone receptor expression in the human vocal fold subunits. Journal of Voice 31(4):476–482. https://doi.org/10.1016/j.jvoice.2016.11.005

29. Klasmeyer, G., & Sendlmeier, W. F. (2000). Voice and emotional states. [In:] Kent. R.D. & M. J. Ball (Eds.), Voice Quality Measurement (pp. 339–357). Singular Publishing Group.

30. Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. Journal of Experimental Social Psychology, 38(6), 618-625. https://doi.org/10.1016/S0022-1031(02)00510-3

31. Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. The Journal of the Acoustical Society of America, 64(4), 1027-1035. https://doi.org/10.1121/1.382086

32. Li, X., Wen, Y., Yang, M., Wang, J., Singh, R., & Raj, B. (2023). Rethinking Voice-Face Correlation: A Geometry View. In *Proceedings of the 31st ACM International Conference on Multimedia* (pp. 2458-2467). https://doi.org/10.1145/3581783.3611779

33. Lucas, T., Hatfield, D., & Henneberg, M. (2023). A morphological comparison between a death mask of the American Prophet Joseph Smith and a photograph likely to depict him. Anthropological Review, 85(4), 1–13. https://doi.org/10.18778/1898-6773.85.4.01

34. Macari, A. T., Karam, I. A., Tabri, D., Sarieddine, D., & Hamdan, A. L. (2015). Formants frequency and dispersion in relation to the length and projection of the upper and lower jaws. Journal of Voice, 29(1), 83–90. https://doi.org/10.1016/j.jvoice.2014.05.011

35. Macari, A. T., Karam, I. A., Ziade, G., Tabri, D., Sarieddine, D., Alam, E. S., & Hamdan, A. L. (2017). Association between facial length and width and fundamental frequency. Journal of Voice, 31(4), 410-415. https://doi.org/10.1016/j.jvoice.2016.12.001

36. Martin R. Lehrbuch der Anthropologie in systematischer Darstellung, Jena. 1914.

37. Marunick, M. T., & Menaldi, C. J. (2000). Maxillary dental arch form related to voice classification: a pilot study. Journal of Voice, 14(1), 82-91. https://doi.org/10.1016/S0892-1997(00)80097-1

38. Mavica, L. W., & Barenholtz, E. (2013). Matching voice and face identity from static images. Journal of Experimental Psychology: Human Perception and Performance, 39(2), 307-312. https://doi.org/10.1167/12.9.1023

39. Mercer, E., & Lowell, S. Y. (2020). The low mandible maneuver: Preliminary study of its effects on aerodynamic and acoustic measures. Journal of Voice, 34(4), 645-e1. https://doi.org/10.1016/j.jvoice.2018.12.005

40. Moreira, T. D. C., Gadenz, C., Figueiró, L. R., Capobianco, D. M., Cunha, K., Ferigolo, M., Barros, H. M. T., & Cassol, M. (2015). Substance use, voice changes and quality of life in licit and illicit drug users. Revista CEFAC, 17, 374-384. https://doi.org/10.1590/1982-021620156714

41. Newman, S. R., Butler, J., Hammond, E. H., & Gray, S. D. (2000). Preliminary report on hormone receptors in the human vocal fold. Journal of Voice, 14(1), 72-81. https://doi.org/10.1016/S0892-1997(00)80096-X

42. Ning, H., Zheng, X., Lu, X., & Yuan, Y. (2021). Disentangled representation learning for cross-modal biometric matching. IEEE Transactions on Multimedia, 24, 1763-1774. https://doi.org/10.1109/TMM.2021.3071243

43. Oh, T. H., Dekel, T., Kim, C., Mosseri, I., Freeman, W. T., Rubinstein, M., & Matusik, W. (2019). Speech2face: Learning the face behind a voice. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7539-7548). https://doi.org/10.1109/cvpr.2019.00772

44. O'Toole, A., Weimer, S., Dunlop, J., Barwick, R., Ayyad, J., & Phillips, J. (2010). Recognizing people from dynamic video: Dissecting identity information with a fusion approach. Journal of Vision, 10(7), 643-643. https://doi.org/10.1167/10.7.643

45. Pawelec, Ł. P., Graja, K., & Lipowicz, A. (2022). Vocal indicators of size, shape and body composition in Polish men. Journal of Voice, 36(6), 878-e9. https://doi.org/10.1016/j.jvoice.2020.09.011

46. Pawelec, Łukasz, Kierczak, K., & Lipowicz, A. (2023). Assessment of the obesity based on voice perception. Anthropological Review, 85(4), 43–60. https://doi.org/10.18778/1898-6773.85.4.04

47. Perini, T. A., Oliveira, G. L. D., Ornellas, J. D. S., & Oliveira, F. P. D. (2005). Technical error of measurement in anthropometry. Revista Brasileira de Medicina do Esporte, 11, 81-85. https://doi.org/10.1590/S1517-86922005000100009

48. Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C. & Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. Animal Behaviour, 95, 89–99. https://doi.org/10.1016/j.anbehav.2014.06.011

49. Pisanski, K., Jones, B. C., Fink, B., O'Connor, J. J., DeBruine, L. M., Röder, S., & Feinberg, D. R. (2016). Voice parameters predict sex-specific body morphology in men

and women. Animal Behaviour, 112, 13-22. https://doi.org/10.1016/j.anbehav.2015.11.008

50. Raj, A., Gupta, B., Chowdhury, A., & Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. Journal of Voice, 24(3), 363-368. https://doi.org/10.1016/j.jvoice.2008.10.005

51. Reinheimer, D. M., Andrade, B. M., Nascimento, J. K., Fonte, J. B., Araújo, I. M., Martins-Filho, P. R., Salvatori, R., Valen¸ca, E. H. O., Oliveira, A. H. A., Aguiar-Oliveira, M. H. & Oliveira-Neto, L. A. (2021). Formant frequencies, cephalometric measures, and pharyngeal airway width in adults with congenital, isolated, and untreated growth hormone deficiency. Journal of Voice, 35(1), 61-68. https://doi.org/10.1016/j.jvoice.2019.04.014

52. Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. The Journal of the Acoustical Society of America, 117(2), 944-955. https://doi.org/10.1121/1.1848011

53. Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. Psychological Science, 24(11), 2235-2243. https://doi.org/10.1177/0956797613492986

54. Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. Journal of Experimental Psychology: Human Perception and Performance, 38(5), 1098-1104. https://doi.org/10.1037/a0028584

55. Roers, F., Mürbe, D., & Sundberg, J. (2009). Voice classification and vocal tract of singers: a study of x-ray images and morphology. The Journal of the Acoustical Society of America, 125(1), 503-512. https://doi.org/10.1121/1.3026326

56. Rojas, S., Kefalianos, E., & Vogel, A. (2020). How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. Journal of Speech, Language, and Hearing Research, 63(2), 533-551.

57. Rothkrantz, L. J., Wiggers, P., Van Wees, J. W. A., & van Vark, R. J. (2004). Voice stress analysis. [In:] Text, Speech and Dialogue: 7th International Conference, TSD 2004, Brno, Czech Republic, September 8-11, 2004. Proceedings 7 (pp. 449-456). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30120-2_57

58. Sondhi, S., Khan, M., Vijay, R., & K. Salhan, A. (2015). Vocal indicators of emotional stress. International Journal of Computer Applications, 122(15), 38–43. https://doi.org/10.5120/21780-5056

59. Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A., Kowal, M., Pisanski, K. (2019). Voice of authority: professionals lower their vocal frequencies when giving expert advice. Journal of Nonverbal Behavior, 43(2), 257–269. https://doi.org/10.1007/s10919-019-00307-0

60. Story, B. H., Titze, I. R., & Hoffman, E. A. (2001). The relationship of vocal tract shape to three voice qualities. The Journal of the Acoustical Society of America, 109(4), 1651-1667. https://doi.org/10.1121/1.1352085

61. Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis – jitter, shimmer and HNR parameters. Procedia Technology, 9, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

62. Titze, I. R. (1994). Fluctuations and perturbations in vocal output. Principles of Voice Production, 209-306.

63. Titze, I. R. (2011). Vocal fold mass is not a useful quantity for describing F0 in vocalization. Journal of Speech Language and Hearing Research 54(2), 520-522. https://doi.org/10.1044/1092-4388(2010/09-0284)

64. Voelter, C., Kleinsasser, N., Joa, P., Nowack, I., Martinez, R., Hagen, R., & Voelker, H. U. (2008). Detection of hormone receptors in the human vocal fold. European Archives of Oto-rhino-laryngology, 265, 1239-1244. https://doi.org/10.1007/s00405-008-0632-x

65. Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results. International Journal of Pediatric Otorhinolaryngology, 49(3), 197-206. https://doi.org/10.1016/S0165-5876(99)00208-6

66. Wen, P., Xu, Q., Jiang, Y., Yang, Z., He, Y., & Huang, Q. (2021). Seeking the shape of sound: An adaptive framework for learning voice-face association. [In:] Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16347-16356). https://doi.org/10.1109/CVPR46437.2021.01608

67. Wood, S. (1986). The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. The Journal of the Acoustical Society of America, 80(2), 391-401. https://doi.org/10.1121/1.394090

68. Wu, C. Y., Hsu, C. C., & Neumann, U. (2022). Cross-modal perceptionist: can face geometry be gleaned from voices? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10452-10461). https://doi.org/10.48550/arXiv.2203.09824

69. Wyganowska-Świątkowska, M., Kowalkowska, I., Mehr, K., & Dąbrowski, M. (2013). An anthropometric analysis of the head and face in vocal students. Folia Phoniatrica et Logopaedica, 65(3), 136-142. https://doi.org/10.1159/000354939

70. Young, A. W., & Bruce, V. (2011). Understanding person perception. British Journal of Psychology, 102(4), 959-974.

71. Zheng, A., Hu, M., Jiang, B., Huang, Y., Yan, Y., & Luo, B. (2021). Adversarial-metric learning for audio-visual cross-modal matching. IEEE Transactions on Multimedia, 24, 338-351. https://doi.org/10.1109/TMM.2021.3050089