

Research Paper

Influence of Facial, Head, and Neck Dimensions on Vocal Acoustic Parameters in Polish Speakers

Łukasz PAWELEC¹, Kamila SŁOWIK, Anna LIPOWICZ*¹*Department of Anthropology, Wrocław University of Environmental and Life Sciences
Wrocław, Poland**Corresponding Author: anna.lipowicz@upwr.edu.pl*Received January 7, 2025; revised October 4, 2025; accepted January 2, 2026;
available online January 29, 2026; version of record March 17, 2026; published issue March 27, 2026.*

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 aims to determine the relationships between face dimensions and acoustic parameters in both sexes and examines 111 adult participants (30 males). Each participant undergoes voice recording, which includes five sustained vowels, along with anthropometric measurements of the neck, head, and face regions. Comparisons between voice parameters and the head, face, and neck regions are conducted employing Pearson's correlation coefficients (r) and a multiple linear regression model. The results reveal significant relationships between head, neck, face dimensions and acoustic parameters in both sexes. Males with higher noses, greater head circumferences, and wider faces tend to have lower formants and more stable voices. Females with larger head circumferences had lower formant values, and those with greater neck circumferences tend to have more stable voices. Also, females with increased nose height have a lower fourth formant (F4). Moreover, females with wider faces, noses, and jaws tend to have less rough voices (lower jitter) and longer maximum phonation time (MPT). These findings may be useful for scientists and law enforcement authorities in creating algorithms that build face models based on voice signals.

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

Copyright © 2026 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are known associations between vocal acoustics and body dimensions (BRUCKERT *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), as well as 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 on the relationship between vocal tract morphology and laryngeal and vocal fold size, as well as 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, e.g., 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 on 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 mea-

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 jaw width (go-go), face width (zy-zy), maxilla width (j-j), or mandibular length (co-gn/co-me), were reported. 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 (KOGELSCHATZ, BARENHOLTZ, 2013) or dynamic facial images (KAMACHI *et al.*, 2003).

Some studies examined relationships between head and neck circumferences and vocal characteristics, but their findings are ambiguous – some of them showed a lack of such relationships (EVANS *et al.*, 2006; RENDALL *et al.*, 2005), whereas others revealed a modest inverse association between neck circumference and voice pitch ($r \sim -0.3$; PAWELEC *et al.*, 2022) or stronger associations between the first three formants (F1-F3) and head circumference ($r \sim -0.6$ to 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 ($r \sim -0.22$ and $r \sim -0.25$, respectively) and inversely with the first formant (F1) and positively with the third formant (F3) in females ($r \sim -0.31$ and $r \sim 0.27$, respectively). MACARI *et al.* (2017) observed an inverse association between mandibular width and vocal pitch (F0) or habitual pitch (HP) in female participants ($r \sim -0.35$ and $r \sim -0.39$, respectively) and inverse correlations between F0 or HP and maxillary width in male participants ($r \sim -0.57$ and $r \sim -0.66$, respectively). This study also revealed negative correlations between face width and HP in males and females ($r \sim -0.54$ and $r \sim -0.35$, respectively). The reason why such relationships between head/neck anatomical structures and voice parameters may be observed is 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); therefore, 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 aims to investigate the interrelationship between head, face, and neck dimensions and various acoustic parameters. The approach used two voice stimuli: five sustained vowels and a short sentence without emotional overtones, rather than only sustained vowel phonation as used in similar studies (BOMMARITO *et al.*, 2019; MACARI *et al.*, 2015; 2017; REINHEIMER *et al.*, 2021). Moreover, anthropometric measurements were used to determine the facial dimensions of live persons using calipers, and the 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 has attempted 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 these studies used some algorithms to match voice and face without identifying the strongest determinants of acoustic parameters, and their findings are inconsistent. Therefore, the current study focuses on identifying head and facial features most strongly associated with voice parameters and determining the direction of these relationships. The relationships revealed could help future researches 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

The study participants included 135 participants (40 males) from Wrocław, Lower Silesia, Poland. The study group consisted of students from the Wrocław University of Environmental and Life Sciences, Faculty of Biology and Animal Science, and adult inhabitants of Wrocław invited to the research. All volunteers were examined at the

same time of day (9 AM to 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), and questions on all possible factors that could affect their acoustic parameters, including 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 singer, 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 receptors of the 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 assigning each participant an individual anonymous code. This study was approved by the Bioethics Committee at the Wrocław Medical University (consent number: KB – 25/2021).

2.2. Voice recording and analysis procedure

The voice recording of each participant was conducted under the same standardized conditions, and the recorded material consisted of five vowels (/a:/, /ɛ:/, /i:/, /ɔ:/, /u:/) sustained for three seconds, with a one-second break between them. This method is one of the most commonly used in such studies (e.g., PISANSKI *et al.*, 2014), although other variants are also used, 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 SM58 SE dynamic cardioid microphone (bandwidth 50 Hz to 15 kHz) connected to an IMG Stageline MPA-202 amplifier and the sound card from a Dell Latitude E6400 computer. Each participant was recorded while positioned in front of a microphone mounted on a height-adjustable 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 recording conditions was achieved using a Mozos Mshield acoustic cabin. The sampling frequency was the same for all recordings and was set to 44.1 kHz (16-bit resolution), and all tracks were saved as uncompressed mono format (.wav). The Benetech GM1351 (Benetech Poland) sound level meter indicated that the background noise level of the recording room used was ~38 dB.

All recorded soundtracks were subsequently analyzed using Praat software version 3.9.2. (BOERSMA, WEENINK, 2019). In the first instance, each vowel was analyzed separately. The middle (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, including fundamental frequency F0 [Hz], defined as the lowest and strongest harmonic produced by vocal fold vibration and perceived as vocal pitch (mean F0, median F0, standard deviation (SD)-F0, min F0, and max F0), jitter [%], defined as the degree of variation in sound wave frequency from period to period (LOCAL, RAP, PPQ5, DDP), shimmer [%], defined as the degree of the amplitude variation of the acoustic wave from period to period (LOCAL, APQ3, APQ5, APG11, DDA), and mean harmonics-to-noise ratio (HNR) [dB], an indicator of the relationship between harmonic and noise components in the voice (TEIXEIRA *et al.*, 2013; TITZE, 1994). Other measurements included formant frequencies (formants: F1–F4 [Hz]), which are formed by filtering F0 in the supralaryngeal vocal tract (FANT, 1960), and their derivatives, including formant number F_n [Hz], formant position P_f [Z], formant spacing ΔF [Hz], and formant dispersion D_f [Hz]. Apparent VTL [cm] estimated from formants values, and voice intensity (loudness [dB]) was also calculated (see the appendix of (PISANSKI *et al.*, 2014)). Final values of acoustic parameters were calculated as the mean across all five vowels. Additionally, mean MPT [s] was measured as the maximum duration of sustaining the vowel /a:/ across 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 set to 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 greater than 88% ($r = 0.94$, $p < 0.001$). 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 cm 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 opisthocranium (op), the most posterior and inferior point on the occipital bone. Facial measurements were taken using a sliding caliper and a 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-alar (al-al), upper lip height subnasale-stomion (sn-sto), lower lip height stomion-supramentale (sto-sm), total lip vermilion height labiale superius-labiale inferius (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), as well as knowledge about relationship between head/face/neck dimensions and vocal tract parameters (FITCH, GIEDD, 1999), the above-mentioned measurements were selected for analyses. The rationale for choosing these head/neck measurements was: a) their strong association with vocal tract dimensions (i.e., oral cavity), and b) their use by other authors addressing similar research topics.

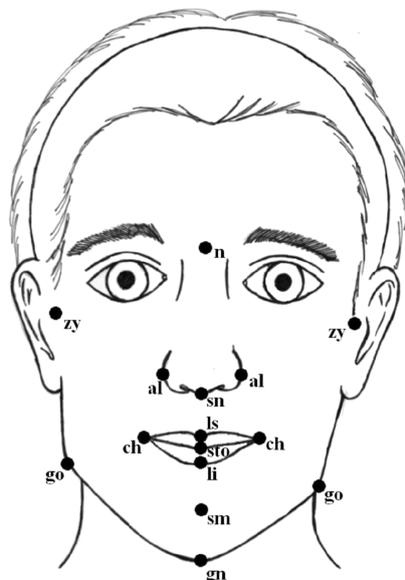


Fig. 1. Facial anthropometric landmarks used for measurements (taken from (GRAJA, KRÓL, 2022) and modified by the authors).

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 by PERINI *et al.* (2005). Both absolute and relative [%] errors were calculated based on the following equations:

$$\text{Absolute TEM} = \sqrt{\frac{\sum d_i^2}{2n}}, \quad (1)$$

where $\sum d_i^2$ is the sum of squared differences between the first and second measurements, n is the number of measured participants, and i denotes the number of measurements;

$$\text{Relative TEM} = \frac{\text{Absolute TEM}}{\text{VAV}} * 100 \%, \quad (2)$$

where VAV is the variable average value (the arithmetic mean calculated based on average values obtained from the two measurements).

2.4. Statistical methods

All analyses were performed using Statistica 13.5 software (1984–2017, TIBCO Software Inc, Palo Alto, California, USA). The 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, an independent sample t -test was applied, and the measures of central tendency and variability were reported as mean \pm SD. When the distribution was not Gaussian in at least one studied group, the Mann–Whitney U -test was applied, and median and lower/upper quartiles (Q1–Q3) were reported. In both cases, the range (min–max) was also shown. The relationship between neck and head measurements and voice parameters were assessed using Pearson’s partial linear correlation, due to the continuous character of the variables, with age and body height included as confounders because of their known associations with 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 greatest impact on voice parameters in males and females, analyzed separately. Several models were created for both sexes, with a single voice parameter as the dependent variable in each model (one model used one voice parameter as a dependent variable). The acoustic parameters included mean F0 (pitch), jitter parameters, shimmer parameters, HNR, intensity, P_f , ΔF , D_f , apparent VTL, and MPT. For males, the explanatory variables (predictors) included head circumference and nose length (n-sn), while for females, three predictors were used: lip vermilion height (ls-li), mandible breadth (go-go), face height (n-gn). The independent variables used in the regression models were selected based on the highest correlation coefficients (r) between voice parameters and anthropometric features for each sex. Collinearity among predictors was tested by correlation coefficients, and redundant variables were excluded from the models. The highest allowed collinearity between predictors was $r \leq 0.3$. In each model, age and body height were included as a confounders. For each regression model, the adjusted coefficient of determination (R^2) was also computed. Only models with significant predictors are reported. 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 from 18 to 65 years), with males being on average five years older than females. Moreover, males had greater values of body height, body mass, neck circumference, head circumference, face width, and most head measurements, except for total lip vermilion height (ls-li; Table 1). The intra-evaluator TEM-absolute [cm or mm] and relative [%] for head and face dimensions were as follows: 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 PERINI *et al.* (2005), the obtained relative TEM values are below 10 % and are therefore considered acceptable.

Table 1. Physical characteristics of study participants.

Feature	Males ($n = 30$)*	Range	Females ($n = 81$)*	Range	p -value**
General data					
Age [years]	27.0 [22.0; 36.0]	20–65	22.0 [22.0; 25.5]	18–64	0.046
Body height [cm]	180.5 \pm 6.5	168.1–197	165.6 \pm 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 \pm 1.5	54–60	55.2 \pm 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 \pm 4.1	5–23	15.1 \pm 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 \pm SD for t -Student test, or median (Q1–Q3) for Mann–Whitney test.

** Significance of male-female differences.

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

Acoustic parameter	Males ($n = 30$)*	Range	Females ($n = 81$)*	Range	p -value**
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
SD [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]	163.0–285.2	<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 HNR [dB]	18.1 \pm 3.3	11.5–23.9	22.4 \pm 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 \pm 74.5	463.4–801.9	590.3 \pm 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 \pm 162.3	2393.9–3079.8	2901.0 \pm 144.5	2561.1–3194.6	<0.001
F4 [Hz]	3730.3 \pm 163.5	3425.6–4072.7	4012.0 \pm 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 \pm 80.2	965.8–1270.9	1137.3 \pm 52.6	1018.7–1273.6	0.030
D_f [Hz]	1053.3 \pm 50.7	962.0–1154.4	1140.6 \pm 68.4	986.5–1324.4	<0.001
Apparent VTL [cm]	15.3 \pm 1.1	13.3–17.5	14.8 \pm 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

The mean pitch (F0) computed for male participants was 110.0 Hz, while for female participants, it was 207.8 Hz. Also, formants, their derivative values, and HNR values were lower in males. On the other hand, instability (jitter and shimmer), intensity, and apparent VTL were higher in males. MPT values were approximately twice as long in males compared to females (Table 2). Most head/face/neck measurements, as well as acoustics parameters, significantly differed between the sexes (Table 1 and Table 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 parameter group, indicating that a larger head circumference is associated with greater amplitude in the sound wave from period to period (i.e., a less stable or more hoarse voice), independent of age and body height. Additionally, males with larger head circumferences had lower HNR values (indicating less stable voice quality), although this relationship did not reach statistical significance. Males with greater nose height (n-sn) had lower values of the second formant (F2, $r = -0.43$, $p = 0.023$) and higher values of apparent VTL ($r = 0.41$, $p = 0.03$; Fig. 2). In contrast, males with wider faces did not show significant changes in HNR ($r = -0.36$, ns) or in the third formant (F3, $r = 0.28$, ns). Also, male participants with increased nose height (n-sn) had shorter MPT, though this correlation was not statistically significant ($r = -0.36$, $0.05 < p < 0.1$; Table 3).

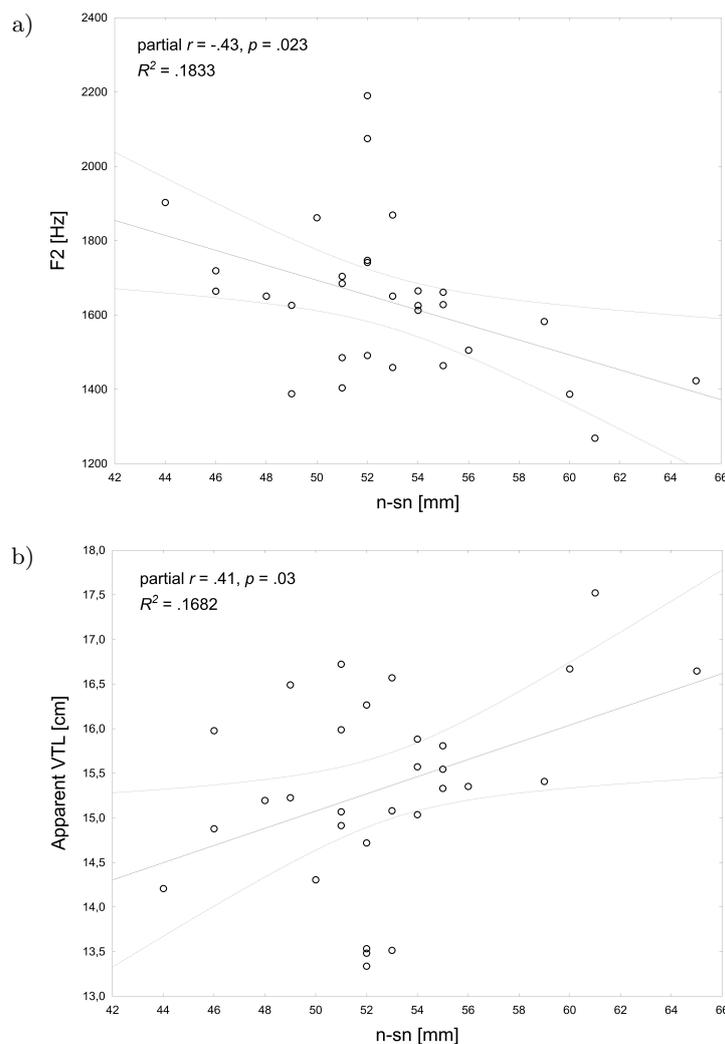


Fig. 2. Nose height (n-sn) and voice formants (from vowels) partial correlations controlling for age and body height as confounders: a) second formant (F2): $r = -0.43$, $p = 0.023$, b) apparent VTL: $r = 0.41$, $p = 0.03$ (male participants).

Table 3. Head/neck circumferences and face measurements vs. voice parameters from vowels.

Pearson's correlation coefficients for males ($n = 30$) and females ($n = 81$);
 $p < 0.05$ results are written in **bold**, ^b – borderline significance ($0.05 < p < 0.06$).

Acoustic parameter	Sex (male/female)	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	-0.02	-0.03	0.12	0.15	0.27	-0.21	0.03	-0.25	0.008	0.12	0.15
	F	0.06	0.08	-0.04	-0.01	-0.10	0.03	-0.20	-0.09	0.08	0.04	0.08
Jitter [%]												
LOCAL	M	0.15	0.12	-0.05	0.11	-0.15	-0.02	0.18	-0.35	0.10	0.21	0.20
	F	-0.10	-0.05	-0.07	-0.09	0.06	-0.10	0.09	0.06	0.03	-0.12	-0.14
RAP	M	0.23	0.31	0.05	-0.01	-0.07	-0.02	0.08	-0.16	0.11	0.31	0.45
	F	-0.09	-0.001	-0.13	-0.13	0.03	-0.10	0.11	0.04	0.05	-0.14	-0.19
PPQ5	M	0.23	0.11	0.04	0.07	-0.10	0.06	0.10	-0.20	0.13	0.18	0.21
	F	-0.10	-0.08	-0.08	-0.04	0.03	-0.08	0.04	0.08	0.03	-0.22	-0.19
DDP	M	0.23	0.31	0.05	-0.006	-0.007	-0.002	0.12	-0.06	0.01	0.001	0.31
	F	-0.04	0.04	-0.12	-0.11	0.04	-0.07	0.13	0.06	0.10	-0.23	-0.15
Shimmer [%]												
LOCAL	M	0.41	0.21	-0.11	0.08	0.04	0.05	0.03	-0.18	0.36^b	0.20	0.21
	F	-0.17	-0.27	0.12	0.14	-0.01	0.17	0.01	0.29	-0.03	0.007	-0.22^b
APQ3	M	0.43	0.26	-0.02	0.03	0.13	0.03	0.06	-0.21	0.38	0.26	0.33
	F	-0.18	-0.27	0.10	0.07	-0.03	0.18	0.02	0.25	-0.04	-0.03	-0.22^b
APQ5	M	0.42	0.22	-0.006	0.01	0.19	0.01	0.04	-0.19	0.11	0.23	0.24
	F	-0.01	-0.25	0.11	0.09	0.07	0.19	0.03	0.29	-0.01	0.03	-0.20
APQ11	M	0.39	0.22	-0.19	0.01	0.01	0.09	-0.02	-0.21	0.38	0.24	0.14
	F	-0.14	-0.25	0.16	0.12	-0.11	0.15	0.01	0.32	0.002	0.06	-0.14
DDA	M	0.43	0.26	0.02	0.03	0.13	0.03	0.06	-0.21	0.21	0.26	0.33
	F	-0.18	-0.27	0.10	0.07	-0.03	0.18	0.02	0.25	-0.04	-0.03	-0.22^b
HNR [dB]	M	-0.32	-0.40	0.03	0.08	-0.19	-0.07	0.03	0.26	-0.29	-0.36	-0.28
	F	0.12	0.17	-0.19	-0.03	-0.07	-0.14	0.04	-0.13	0.03	0.13	0.13
Intensity [dB]	M	-0.17	-0.08	-0.005	-0.15	0.18	-0.15	0.01	-0.21	-0.29	-0.12	-0.16
	F	0.06	0.08	-0.46	-0.21	0.10	-0.11	0.11	-0.11	-0.07	-0.13	0.03
Formants and their derivatives												
F1 [Hz]	M	0.003	0.11	-0.24	-0.25	0.22	-0.05	0.06	-0.21	-0.06	-0.06	-0.08
	F	-0.11	-0.16	0.02	-0.05	-0.07	0.14	-0.07	-0.08	-0.06	-0.08	-0.10
F2 [Hz]	M	0.17	0.25	-0.14	-0.43	0.30	0.13	0.02	-0.26	0.01	0.18	0.10
	F	-0.15	-0.04	0.07	-0.21	-0.006	-0.05	-0.09	-0.15	-0.005	-0.07	0.03
F3 [Hz]	M	0.03	0.19	-0.13	-0.21	0.07	-0.16	-0.26	-0.26	0.12	0.28	0.24
	F	-0.10	-0.13	0.03	0.11	0.13	-0.06	0.11	-0.05	0.10	-0.12	0.01
F4 [Hz]	M	0.05	0.17	0.08	-0.09	0.15	0.09	-0.06	0.03	0.07	0.13	0.01
	F	-0.003	0.08	-0.16	-0.24	0.11	0.03	-0.20	-0.17	-0.08	-0.08	0.12
P_f [Z]	M	0.10	0.25	-0.12	-0.33	0.24	0.03	-0.09	-0.27	0.06	0.21	0.12
	F	-0.11	-0.05	-0.05	-0.15	0.10	0.002	-0.13	-0.13	-0.01	-0.13	0.08
ΔF [Hz]	M	0.09	0.22	-0.19	-0.35	0.27	0.01	-0.03	-0.29	0.01	0.13	0.06
	F	-0.15	-0.13	0.003	-0.14	0.02	0.06	-0.10	-0.10	0.05	-0.12	-0.07
D_f [Hz]	M	0.05	0.12	0.20	0.03	0.04	0.12	0.25	0.14	0.10	0.17	0.05
	F	0.03	0.12	-0.15	-0.17	0.12	-0.01	-0.07	-0.05	-0.05	-0.05	0.15
Apparent VTL [cm]	M	-0.11	-0.25	0.22	0.41	-0.28	-0.04	0.02	0.27	-0.02	-0.16	-0.08
	F	0.14	0.12	0.01	0.14	-0.03	-0.07	0.12	0.14	0.04	0.14	0.07
MPT [s]	M	0.08	-0.006	-0.03	-0.36	0.12	0.001	-0.35	-0.08	0.27	0.22	0.12
	F	0.08	0.19	0.17	0.08	-0.04	0.07	0.04	-0.03	-0.07	-0.07	0.25

For female participants, more significant associations between head and face dimensions and acoustic parameters extracted from vowels were observed (Table 3). Female participants with larger neck circumferences had lower values of shimmer parameters (APQ3, DDA: $r = -0.27$, $p = 0.018$, and APQ5, APQ11: $r = -0.25$, $p = 0.032$). Furthermore, females with higher face height (n-gn) had lower voice intensity (loudness: $r = -0.46$, $p < 0.001$).

Additionally, a higher nose height (n-sn), lower fourth formant (F4), and longer apparent VTL showed a significant positive relationship with lip vermilion height (ls-li) and voice parameters. A higher value of this characteristic

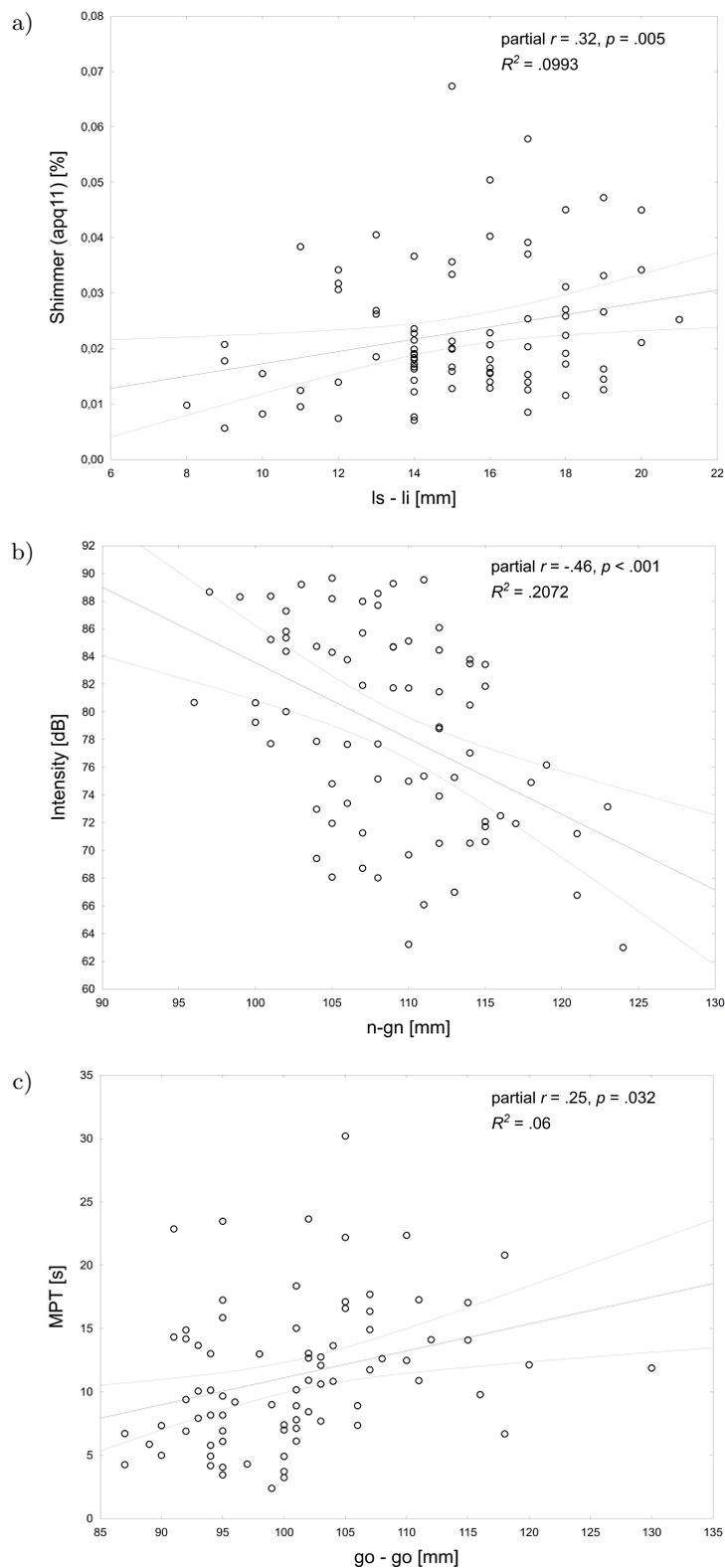


Fig. 3. Partial linear correlations controlling for age and body height as confounders in female participants: a) lip vermilion height (ls-li) and shimmer (APQ11): $r = 0.32$, $p = 0.005$, b) face height (n-gn) vs. voice intensity (loudness): $r = -0.46$, $p < 0.001$, c) jaw width (go-go) and MPT: $r = 0.25$, $p = 0.032$.

was positively correlated with shimmer parameters, indicating that greater lip vermilion was associated with an unstable (hoarse) voice. This was related to a greater face width (zy-zy) and lower jitter values (RAP, PPQ5, DDP). Also, jaw width (go-go) showed similar trends with jitter parameters, although 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 aforementioned correlations were weak (~ 0.2 – 0.3) as it was expected, but some of them (i.e., nose height (n-sn) vs. F2 showed a moderate effect size for male 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 between head circumference and shimmer (LOCAL), meaning that males with larger head circumferences had more unstable (hoarse) voices ($\beta = 0.41$, $t = 2.25$, $p = 0.034$; see Table 4, model 1). The second model revealed a significant correlation between nose height (n-sn) and apparent VTL, indicating that males with higher noses had longer estimated vocal tracts ($\beta = 0.42$, $t = 2.28$, $p = 0.031$; see Table 4, model 2).

Table 4. Multiple general regression models (GRMs) 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 -value
Males	1. Dependent variable: shimmer (LOCAL) [%] $R^2_{Adj.} = 13.29\%$, $\varepsilon = 0.017$ $F = 3.22$, $p < 0.06$	Intercept	–	–	–1.73	0.097
		Head circumferences [cm]	0.41	0.18	2.25	0.034
		n-sn [mm]	0.07	0.19	0.38	0.704
		Age [years]	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] $R^2_{Adj.} = 12.42\%$, $\varepsilon = 0.99$ $F = 3.06$, $p < 0.06$	Intercept	–	–	1.15	0.260
		Head circumferences [cm]	–0.12	0.18	–0.65	0.522
		n-sn [mm]	0.42	0.18	2.28	0.031
		Age [years]	–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) [%] $R^2_{Adj.} = 13.45\%$, $\varepsilon = 0.011$ $F = 3.43$, $p < 0.008$	Intercept	–	–	–2.52	0.014
		ls-li [mm]	2.99	0.004	2.99	0.004
		go-go [mm]	–1.52	0.132	–1.52	0.132
		n-gn [mm]	1.84	0.070	1.84	0.070
		Age [years]	0.33	0.13	2.44	0.017
		Body height [cm]	0.31	0.12	2.71	0.008
	4. Dependent variable: intensity [dB] $R^2_{Adj.} = 0.16\%$, $\varepsilon = 7.15$ $F = 1.02$, $p < 0.410$	Intercept	–	–	4.44	< 0.001
		ls-li [mm]	–0.13	0.12	–1.07	0.288
		go-go [mm]	0.06	0.13	0.48	0.632
		n-gn [mm]	–0.23	0.12	–1.96	0.054^b
		Age [years]	–0.20	0.14	–1.37	0.175
		Body height [cm]	–0.09	0.13	–0.73	0.467
	5. Dependent variable: MPT [s] $R^2_{Adj.} = 8.59\%$, $\varepsilon = 5.52$ $F = 2.47$, $p < 0.040$	Intercept	–	–	–2.18	0.033
		ls-li [mm]	–0.02	0.12	–0.16	0.875
		go-go [mm]	0.24	0.12	1.98	0.052^b
n-gn [mm]		0.14	0.11	1.23	0.221	
Age [years]		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^2 of the model, ε – standardized random error, t – t -statistic of β showing predictor significance, F – Fisher test value showing overall model significance. The results significant on the $p < 0.05$ level are written in **bold**, ^b – borderline significance ($0.05 < p < 0.06$).

For females, the first multiple regression model revealed that lip vermilion height (ls-li) significantly correlated with shimmer (APQ11), indicating that a larger ls-li ($\beta = 0.35$, $t = 2.99$, $p = 0.004$) was associated with a more hoarse voice ($\beta = 0.35$, $t = 2.99$, $p = 0.004$; see Table 4, Model 3). The second model showed a borderline

significance negative correlation between face height (n-gn) and voice intensity (loudness), which means that females 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 a borderline significance correlation between jaw width (go-go) and MPT. This associations means that women with wider jaws had longer maximum time of phonation ($\beta = 0.24$, $t = 1.98$, $p = 0.052$, see Table 4, Model 5).

4. Discussion

Significant relationships between head, neck, and face dimensions and voice parameters were observed for both sexes. The strongest relationships for males occurred between shimmer parameters and head circumference, F2, nose height (n-sn) and apparent VTL, as well as jitter (RAP) and jaw width (go-go). Thus, males with larger heads exhibited more hoarse (unstable) voices, and those with higher noses were associated with lower values of F2 and longer apparent VTL. Males with wider jaws had greater voice roughness. For females, the strongest connections included: face height (n-gn) and voice intensity, neck circumference and lip vermilion height (ls-li) with shimmer parameters, nose height (n-sn) and F4, and jaw width (go-go) and MPT. These results suggest that females with smaller face heights had higher voice intensity, and those with greater neck circumferences and lower lip height had more stable (less hoarse) voices. Finally, females with longer noses had lower F4, while 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), or their derivatives (P_f or D_f) and head or neck circumferences for males and females, although such relationships were found for shimmer parameters. These results partly confirm EVANS *et al.* (2006), who reported the lack of such relationships for males, and support RENDALL *et al.* (2005), who found no significant connections for both sexes. In contrast, PAWELEC *et al.* (2022) reported significant associations between neck circumference and formants and F0 in males, even after applying multiple regression models.

For females, the strongest relationships based on multiple linear regression models were reported for jaw width and MPT (jaw width had the strongest effect on MPT). Interestingly, no significant associations were found between voice pitch (F0) and head, neck, or 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 (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.

The observed connections between facial 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). Larger individuals also have a bigger larynx and a longer vocal tract (FITCH, GIEDD, 1999), and such anatomical structures affect the voice through longer vocal folds, lower F0, a longer vocal tract, and lower formant frequencies (FITCH, 1997; TITZE, 2011). Therefore, an inverse relationship between body size and voice parameters is observed (BRUCKERT *et al.*, 2006; RENDALL *et al.*, 2005), which has also been noted in subjective assessment of body build conducted by judges (PAWELEC *et al.*, 2023). A positive correlation has also been found between head length (g-op), face height (n-gn), and growth indexes of vocal tract structures during early childhood (VORPERIAN *et al.*, 1999). Thus, the 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, independent of body size and age. Body size (expressed as body height) and chronological age of the participants were controlled as confounders in the partial correlations and regression models; thus, their apparent impact on head/neck/face – voice relationships was reduced.

Relationships between jaw width and acoustic parameters in females may seem difficult to explain at first glance. However, females with greater face width tend to have lower jitter values (i.e., more stable voices). Moreover, females with wider mandibles have lower shimmer values (i.e., less hoarse voice) and greater MPT. In another study, it was also reported that female vocalists had significantly greater facial dimensions and mandible width than a control group of non-vocalists (WYGANOWSKA-ŚWIĄTKOWSKA *et al.*, 2013). These results may suggest a role for facial and jaw morphology in voice production (especially articulation) and speech processes. Larger jaw size is likely associated with stronger development of the muscles attached to the mandible and greater development of the entire stomatognathic apparatus. Consequently, this may affect the biomechanics of these structures during voice production. Some support for this thesis is provided by results showing that nonsingers using a low mandible maneuver (LMM), a technique employed by top-ranked professional singers to enhance vocal output by altering oral, pharyngeal, and laryngeal configurations, demonstrated increased voice intensity (sound-pressure level (SPL)) and lower F1 and F2 values (MERCER, LOWELL, 2020).

The significant negative relationships between face height, nose height, intensity, and formants, as well as the positive correlations between these dimensions and apparent VTL in both males and females, can be explained by the anatomy of the speech apparatus. Facial morphology, such as face and nose height, is associated with the height of oral and nasal cavity and the pharyngeal airway (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). Thus, the aforementioned relationships likely result in an apparent correlation between face and nose height and the formant frequencies of the voice.

No associations between lip vermilion height and formants or their derivatives in either sex were found. This result is puzzling, as some studies 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, 1986). A significant relationship was also reported between 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 F1 and mouth height, but no correlation between F2 and mouth height (ERBER, 1979). In the present study, lip vermilion height was positively correlated 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 additional 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 additional vocal tract structures, such as the frontal sinuses, on voice characteristics. Another limitation is the lack of head and neck imaging, such as X-ray, computed tomography, and magnetic resonance imaging, which would have allowed the true dimensions of the participants' vocal tracts to be determined and their influence on voice acoustic parameters to be highlighted. In addition, the unequal number of male and female participants makes cross-sex comparisons difficult. Some of the relationships presented in this study may have limited practical applicability due to the methodology used – isolated sustained vowels. Correlations with acoustic parameters computed from short sentence or spontaneous speech are likely to have greater ecological validity, as these signals more closely resemble natural conditions. Another limitation of the study is the lack of control for other factors that may influence 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), use of 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 is an evidence 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 relationship between vocal tract structure and function and voice production. Indeed, both facial width and length measurements, in particular, showed significant negative relationships with voice parameters (BOMMARITO *et al.*, 2019; MACARI *et al.*, 2015; 2017; REINHEIMER *et al.*, 2021). Connections were also observed between head and neck circumferences and acoustic parameters (PAWELEC *et al.*, 2022; REINHEIMER *et al.*, 2021). VTL and shape were associated with voice parameters, especially formant frequencies (FITCH, 1997; STORY *et al.*, 2001), as well as body size, height, and weight (FITCH, GIEDD, 1999). Moreover, head and face size and proportions were related to body size, VTL, and shape, indicating that correlations may 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 could potentially be applied by forensic scientists to estimate the facial morphology of offenders based solely on voice recordings (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 meta-analyses of the relationship between facial morphology and voice acoustic parameters. Another promising approach is to use geometric morphometrics, which considers not only the size of the head/face but also their shape through indices describing facial geometry. The results of such studies would enable investigators to identify perpetrators and victims based solely on audio recordings or to evaluate voice changes in people following surgical interventions of the head and face (medical and aesthetic plastic surgery). Recent research indicates some degree of success in estimating facial morphology from the voice signal (LI *et al.*, 2023).

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 Ministry of Science and Higher Education 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

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHORS’ CONTRIBUTIONS

Łukasz Pawelec conceptualized the study and wrote the original draft, performed the analysis, and contributed to data interpretation, investigation, visualization, review, and editing. Kamila Słowik contributed to the investigation, conceptualization, performed the analysis. Anna Lipowicz contributed to conceptualization, wrote the original draft, project administration and supervision. All authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

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

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](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, <http://www.praat.org/> (access: 20.06.2019).
4. BOMMARITO S. *et al.* (2019), Correlation between voice, speech, body and facial types in young adults, *Global Journal of Otolaryngology*, **20**(4): 556041, <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.e11–501.e18, <https://doi.org/10.1016/j.jvoice.2019.09.015>.
6. BRATSTRÖM 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. BRUCKERT L., LIÉNARD J.-S., LACROIX A., KREUTZER M., LEBOUCHER G. (2006), Women use voice parameters to assess men's characteristics, *Proceedings. Biological Sciences*, **273**(1582): 83–89, <https://doi.org/10.1098/rspb.2005.3265>.
8. BUNKER D. (2017), Speech2Face: Reconstructed lip syncing with generative adversarial networks. Data reflexions: Thoughts and Projects.
9. 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>.
10. 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>.
11. CAZACU C.J. *et al.* (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, <https://doi.org/10.47162/RJME.66.2.10>.
12. 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](https://doi.org/10.1016/S0191-8869(98)00272-4).
13. 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>.
14. 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>.
15. 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>.
16. 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>.
17. FANT G. (1960), *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*, 2nd ed., Walter de Gruyter.
18. 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>.
19. 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>.
20. 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](https://doi.org/10.1016/S0095-4470(03)00049-4).

21. 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.105.1.215-220>.
22. GRADDOL D., SWANN J. (1983), Speaking fundamental frequency: Some physical and social correlates, *Language and Speech*, **26**(4): 351–366, <https://doi.org/10.1177/002383098302600403>.
23. GRAJA K., KRÓL B. (2022), *Anthropometry and Anthroposcopy. Lecture Notes* [in Polish: *Antropometria i Antroposkopia. Skrypt do Ćwiczeń*], Unpublished teaching materials of the Division of Anthropology, Wrocław University of Environmental and Life Sciences.
24. HAMDAN A.L. et al. (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>.
25. HAMDAN A.L.H. et al. (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>.
26. JANDOVÁ M., URBANOVÁ P. (2016), The relationship between facial morphology, body measurements and socio-economic factors, *Anthropological Review*, **79**(2): 181–200, <https://doi.org/10.1515/anre-2016-0014>.
27. 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>.
28. 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>.
29. 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, <https://doi.org/10.1109/6ICMI.2002.1167010>.
30. 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>.
31. KLASMEYER G., SENDLMEIER W.F. (2000), Voice and emotional states, [in:] *Voice Quality Measurement*, Kent R.D., Ball M.J. [Eds.], pp. 339–357, Singular Publishing.
32. KOGELSCHATZ L., BARENHOLTZ E. (2013), Matching voice and face identity from static images, *Journal of Vision*, **12**(9): 1023, <https://doi.org/10.1167/12.9.1023>.
33. 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](https://doi.org/10.1016/S0022-1031(02)00510-3).
34. 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>.
35. LI X., WEN Y., YANG M., WANG J., SINGH R., RAJ B. (2023), Rethinking voice-face correlation: A geometry view, [in:] *MM '23: Proceedings of the 31st ACM International Conference on Multimedia*, pp. 2458–2467, <https://doi.org/10.1145/3581783.3611779>.
36. 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>.
37. MACARI A.T. et al. (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>.
38. 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>.
39. MARTIN R. (1914), *A Textbook of Anthropology: A Systematic Presentation* [in German: *Lehrbuch der Anthropologie in Systematischer Darstellung*], Jena, Gustav Fischer.
40. 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](https://doi.org/10.1016/S0892-1997(00)80097-1).
41. 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.e1–645.e9, <https://doi.org/10.1016/j.jvoice.2018.12.005>.

42. MOREIRA T. de C. *et al.* (2015), Substance use, voice changes and quality of life in licit and illicit drug users [in Portuguese], *Revista CEFAC*, **17**(2): 374–384, <https://doi.org/10.1590/1982-021620156714>.
43. 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](https://doi.org/10.1016/S0892-1997(00)80096-X).
44. 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>.
45. OH T.H. *et al.* (2019), Speech2Face: Learning the face behind a voice, [in:] *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7539–7548, <https://doi.org/10.1109/cvpr.2019.00772>.
46. 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, <https://doi.org/10.1167/10.7.643>.
47. 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.e9–878.e22, <https://doi.org/10.1016/j.jvoice.2020.09.011>.
48. PAWELEC Ł., 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>.
49. PERINI T.A., OLIVEIRA G.L. de, ORNELLAS J. dos S., OLIVEIRA F.P. de (2005), Technical error of measurement in anthropometry [in Portuguese], *Revista Brasileira De Medicina Do Esporte*, **11**(1): 81–85, <https://doi.org/10.1590/S1517-86922005000100009>.
50. PISANSKI K. *et al.* (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>.
51. PISANSKI K. *et al.* (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>.
52. RAINE J., PISANSKI K., SIMNER J., REBY D. (2019), Vocal communication of simulated pain, *Bioacoustics*, **28**(5): 404–426, <https://doi.org/10.1080/09524622.2018.1463295>.
53. 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>.
54. REINHEIMER D.M. *et al.* (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>.
55. RENDALL D., KOLLIAS S., NEY C., LLOYD P. (2005), Pitch (F_0) 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>.
56. 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>.
57. 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>.
58. 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>.
59. 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, https://doi.org/10.1044/2019_jslhr-19-00099.
60. ROTHKRANTZ L.J.M., WIGGERS P., VAN WEES J.W.A., VAN VARK R.J. (2004), Voice stress analysis, [in:] *Text, Speech and Dialogue. TSD 2004. Lecture Notes in Computer Science*, Sojka P., Kopeček I., Pala K. [Eds.], Vol. 3206, pp. 449–456, https://doi.org/10.1007/978-3-540-30120-2_57.
61. SONDHI S., KHAN M., VIJAY R., SALHAN A.K. (2015), Vocal indicators of emotional stress, *International Journal of Computer Applications*, **122**(15): 38–43, <https://doi.org/10.5120/21780-5056>.

62. SOROKOWSKI P. et al. (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>.
63. 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>.
64. 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>.
65. TITZE I.R. (1994), Fluctuations and perturbations in vocal output, [in:] *Principles of Voice Production*, Prentice Hall, pp. 209–306.
66. TITZE I.R. (2011), Vocal fold mass is not a useful quantity for describing F_0 in vocalization, *Journal of Speech Language and Hearing Research*, **54**(2): 520–522, [https://doi.org/10.1044/1092-4388\(2010/09-0284\)](https://doi.org/10.1044/1092-4388(2010/09-0284)).
67. VOELTER C. et al. (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>.
68. 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](https://doi.org/10.1016/S0165-5876(99)00208-6).
69. 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:] *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16347–16356, <https://doi.org/10.1109/CVPR46437.2021.01608>.
70. 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>.
71. WU C.Y., HSU C.C., NEUMANN U. (2022), Cross-modal perceptionist: Can face geometry be gleaned from voices?, [in:] *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10452–10461, <https://doi.org/10.1109/CVPR52688.2022.01020>.
72. 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>.
73. YOUNG A.W., BRUCE V. (2011), Understanding person perception, *British Journal of Psychology*, **102**(4): 959–974, <https://doi.org/10.1111/j.2044-8295.2011.02045.x>.
74. 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>.