RESEARCH ARTICLE

No evidence for an association between facial fluctuating asymmetry and vocal attractiveness in men or women

Tobias L. Kordsmeyer[1]* (iD), Yasmin T. K. Thies[1], Omid Ekrami[2] (iD), Julia Stern[1], Christoph Schild[3], Cristina Spoiala[4], Peter Claes[5], Stefan Van Dongen[2] and Lars Penke[1]

[1]Department of Psychology and Leibniz Science Campus, Primate Cognition, University of Goettingen, Gosslerstr. 14, 37073 Goettingen, Germany, [2]Department of Biology, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium, [3]Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen, Denmark, [4]Nivel, Netherlands Institute for Health Services Research, Otterstraat 118, 3513 CR Utrecht, The Netherlands and [5]Department of Electrical Engineering–ESAT & Department of Human Genetics, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
*Corresponding author. E-mail: tob.kor@gmail.com

Abstract

Facial fluctuating asymmetry (FA), presumably a proxy measure of developmental instability, has been proposed to inversely relate to vocal attractiveness, which may convey information on heritable fitness benefits. Using an improved method of measuring facial FA, we sought to replicate two recent studies that showed an inverse correlation of facial FA with vocal attractiveness. In two samples of men ($N = 165$) and women ($N = 157$), we investigated the association of automatically measured facial FA based on 3D face scans with male and female observer-rated attractiveness of voice recordings. No significant associations were found for men or women, also when controlling for facial attractiveness, age, and body mass index. Equivalence tests show that effect sizes were significantly smaller than previous meta-analytic effects, providing robust evidence against a link of facial FA with vocal attractiveness. Thus, our study contradicts earlier findings that vocal attractiveness may signal genetic quality in humans via an association with FA.

Keywords: Facial fluctuating asymmetry; developmental instability; vocal attractiveness; 3D face scans; geometric morphometrics

Media summary: New study showing robust evidence against a significant link of men's and women's facial fluctuating asymmetry with voice attractiveness.

Introduction

Sexual selection is a type of natural selection that favours traits which aid in competition over mates. The two underlying main mechanisms are intersexual attraction of and intrasexual competition for mates (Puts, 2016). The former is based on a preference for opposite-sex individuals providing fitness benefits, including indirect, heritable fitness benefits (Gangestad & Thornhill, 1997; Gangestad et al., 2005). Heritable fitness benefits partly reflect an organism's capacity to buffer against interferences caused by developmental noise (Palmer & Strobeck, 2003). Random variations in developmental factors during ontogeny, detrimental mutations, or external stressors such as diseases, parasites and an inadequate diet impair ideal developmental conditions to some extent (Graham & Özener, 2016; Palmer & Strobeck, 2003). The inability to buffer against these perturbations is referred to as developmental instability (DI), which is assumed to indirectly signal low genetic quality (Van Dongen &

Gangestad, 2011). Since the same genes equally influence development of both sides of the body (Gangestad & Thornhill, 1997; Van Valen, 1962), bilateral features should be expected to show perfect symmetry. However, DI is thought to morphologically manifest in interindividual differences in fluctuating asymmetry (FA). FA appears in bilateral morphological traits for which the differences between left and right sides of the body have a normal distribution with a population mean of zero (Van Valen, 1962). FA is thus considered an indicator of DI, with higher DI resulting in higher FA. In previous studies, FA has been shown to be related to different kinds of outcomes that can be considered as indicators of poor health in both non-human (e.g. probability of survival in the striped dolphin, Pertoldi et al., 2000; ejaculate quality in gazelles, Roldan et al., 1998) and human animals (maternal risk factors, Singh & Rosen, 2001; schizophrenia, Yeo et al., 1999; for a meta-analysis, see Van Dongen & Gangestad, 2011). Concerning heritable fitness benefits, low FA was found to be related to greater mating success in men. For example, men's ($N = 203$) number of self-reported extra-pair copulation (EPC) partners (Gangestad &Thornhill, 1997) and lifetime sexual partners ($N = 60$, Thornhill & Gangestad, 1994) were negatively correlated with their bodily FA scores (but see Kordsmeyer & Penke, 2017 for null-replications of these findings, $N = 284$).

Since FA is assumed to be an indicator of genetic quality via its link with DI, it may also relate to an individual's physical attractiveness. Several studies have shown inverse correlations between FA and visual perceptions of physical attractiveness in humans (bodily FA, Brown et al., 2008, $N = 77$ target participants, $N = 87$ raters; facial FA, Grammer & Thornhill, 1994, $N = 32$ target participants, $N = 96$ raters; Komori et al., 2009, $N = 96$ target participants, $N = 114$ raters; Mogilski & Welling, 2017, $N = 6$ target participants, $N = 504$ raters; Rhodes et al., 1998, $N = 48$ target participants, $N = 64$ raters; but see Jones & Jaeger, 2019 and Simmons et al., 2004 for null-findings on facial FA). Since DI is not perceptible owing to it merely being an underlying mechanism of attractiveness, drawing conclusions about the relationship between the two is problematic. The relationship between FA and DI can be confounded by directional asymmetry (DA), which is assumed to be uninformative of DI and appears in traits where the left–right difference across a population has a normal distribution with a mean deviating from zero (Palmer & Strobeck, 2003). In a study on male facial attractiveness and facial FA, it was shown that horizontal facial FA (HFA, aggregate of the absolute horizontal left–right difference of six distances from a vertical midline based on all bilateral landmarks) and vertical facial FA (VFA, sum of all vertical distances between seven bilateral pairs of landmarks) were inversely related to ratings of attractiveness ($N = 40$ target participants, $N = 79$ raters, Scheib et al., 1999; see also Grammer & Thornhill, 1994). Further studies found FA to be associated with attractiveness perceptions even in the absence of direct visual cues, for example for rated attractiveness of participants' body odour ($N = 77$ target participants, $N = 87$ raters, Gangestad & Thornhill, 1998). Furthermore, vocal qualities have been shown to negatively correlate with different measures of FA (e.g. bodily FA, Abend et al., 2015, $N = 42$ target participants, $N = 103$ raters; Hughes et al., 2002, $N = 106$ target participants, $N = 13$–17 raters; Hughes et al., 2008, $N = 76$ target participants, $N = 101$ raters), suggesting that voices may convey information about DI.

Vocal qualities have also been related to mating success. For both sexes, opposite-sex ratings of vocal attractiveness were related to reported age of first sexual intercourse, number of sexual partners, EPC partners and number of partners by whom they had been chosen as an EPC partner ($N = 96$, Hughes et al., 2004). Vocal characteristics may hence reflect aspects of mate quality and signal information on heritable fitness benefits (Hill & Puts, 2016).

Recent studies provided evidence for associations of vocal attractiveness not only with bodily FA, but also with facial FA. In Abend et al.'s study (2015), the voices of women with more symmetrical bodies and faces were rated as more attractive by male judges ($N = 42$ target participants, $N = 103$ raters). Hill et al. (2017) showed comparable effects across three studies for facial FA. In study 1, men and women's ($N = 325$) vocal attractiveness was rated by unfamiliar participants ($N = 1127$) of the opposite sex. Facial FA was computed as the aggregate of HFA and VFA (based on 2D photographs). Results showed an inverse relationship between facial FA and vocal attractiveness in both men and women. In study 2, in a mixed-sex sample of Hadza hunter–gatherers ($N = 65$), who are thought to have a higher exposure to environmental stressors compared with Western populations

which may influence the development of FA, effects were in the same direction as in study 1, albeit not significant. In a third study, the stimulus set consisted of 2D facial photographs ($N = 79$ males) and 3D face scans for a subsample ($n = 52$). Again, correlations were non-significant but in the predicted direction for both 2D and 3D stimuli. An internal meta-analysis (including their three studies and previously published work from Abend et al., 2015 and Hughes et al., 2002, the latter on bodily FA only), showed a robust negative association between facial FA and vocal attractiveness. Collectively, this suggests that vocal attractiveness may be an indicator of low DI and convey information on heritable fitness benefits. However, non-significant findings (studies 2 and 3 in Hill et al., 2017) call for independent replication to further assess the magnitude and robustness of these associations.

The heterogeneity of results may partly be explained by extant studies differing considerably in methodological respects. Hill et al. (2017) assessed facial FA using measures from both 2D photographs and 3D face scans, whereas most earlier studies used 2D photographs only. Somewhat surprisingly, Hill et al. (2017) observed a null correlation between facial FA measured from 2D photographs and 3D scans, questioning the validity or reliability of either of these methods. Since faces are 3D structures, measuring facial FA from 2D photographs may miss potentially important information from the third dimension. Hence using 3D scans should be the preferred method providing more complete measures of facial FA (Abend et al., 2015; Berssenbrügge et al., 2014; Ekrami et al., 2018; Hill et al., 2017). Moreover, most earlier studies measured FA using simple manual landmarks (e.g. Scheib et al., 1999), for which a diverse array of variants exists (e.g. horizontal and vertical FA, Hill et al., 2017; Scheib et al., 1999; Simmons et al., 2004; comprehensive FA, Penke et al., 2009), and which may miss important information such as dominant features of a face owing to the limited number of landmarks (Ekrami et al., 2018). Geometric morphometrics have been proposed as an improved method capturing greater shape information (e.g. Abend et al., 2015; Claes et al., 2011; Hill et al., 2017; for a review see Mitteroecker & Gunz, 2009). In our study, we combined both favourable approaches using 3D stimuli and utilizing geometric morphometrics. We employed a 3D spatially dense approach for measuring asymmetry, by non-rigidly mapping a symmetrical mask onto the 3D face scans and comparing the faces with their mirrored versions (Claes et al., 2011; Ekrami et al., 2018).

The main goal of our study was to replicate and assess the robustness of the previously reported association between facial FA and vocal attractiveness (Abend et al., 2015; Hill et al., 2017) in a sample of Western men and women, using an improved methodology of automatically measuring facial FA from 3D scans. Based on previous results (Abend et al., 2015; Hill et al., 2017), we hypothesized lower facial FA for males and females with higher rated vocal attractiveness, and that this effect would hold when controlling for facial attractiveness (since facial symmetry may indirectly increase vocal attractiveness via facial attractiveness, for example when individuals with more symmetrical and hence attractive faces speak more confidently and attractively, Hill et al., 2017; Van Dongen & Gangestad, 2011). We further tested for the correlation between 2D and 3D facial FA to examine the convergence of results between these two methods of measuring facial FA.

Methods

Participants

One-hundred and sixty-five men (age: mean (M) = 24.28 years, standard deviation (SD) = 3.25, range 18–34) were recruited as part of a larger study on personality and hormonal reactivity (Kordsmeyer & Penke, 2019), and 157 women (age: M = 23.20 years, SD = 3.45, range 18–34) participated in a larger study on ovulatory cycle effects on, amongst others, mate preferences (Jünger et al., 2018). Owing to sexual dimorphism in faces and voices (e.g. Puts et al., 2012), results were analysed separately for both sexes.

3D facial FA measurement

Participants' faces were 3D-scanned twice for men and overall twelve times for women (three times for each of two sessions during the luteal phase and two sessions during the fertile phase of their ovulatory

cycles, of which the scans from the first luteal sessions were used, see Jünger et al., 2018 and Stern et al., 2019 for details) using a 3dMD face scanner. For each participant one scan was chosen with the most neutral facial expression and standardized head position, while some participants (males: $n = 28$, females: $n = 4$) had to be excluded owing to poor scan quality or interfering facial hair for the males. We used a 3D spatially dense approach to measure FA in the faces (Ekrami et al., 2018). A symmetrical template face comprising 7160 densely located paired vertices was mapped onto each of the faces in the dataset, using a non-rigid iterative closest points algorithm. The reflections of the faces were obtained by reversing the sign of the x-coordinates of the vertices (the x-axis passes through the middle of the face) and then relabelling the vertices with their paired counterparts in order to re-gain compatibility of the original face and its mirror (Klingenberg et al., 2002). The original faces and their reflections were then aligned using a weighted Procrustes algorithm and the total asymmetry for each face was simply calculated by subtracting the mirrored face from the original (see Figure 1 for heat maps of average facial asymmetries). Following the conventional approach (e.g. Palmer & Strobeck, 2003), DA was defined as the average asymmetry of the sample set and FA was obtained by correcting the calculated total asymmetries for DA. For a subset of the men's sample ($n = 45$), facial asymmetry was measured from a second facial scan as described above to examine retest reliability of these automatic measures. These two measures were highly correlated ($r = 0.92$, $p < 0.001$), underlining the reliability of these automatic facial FA measures.

Facial FA from manual landmarks

To be able to directly compare results from these two samples with previously published studies (e.g. Abend et al., 2015; Hill et al., 2017), correlations of various facial FA measures based on landmarks manually placed on 2D facial photographs (for women only) and on 3D face scans (for men and women) with vocal attractiveness are reported in the Supplementary Material (Table S1, for information on methods see Supplementary Material).

Voice recordings

Men's voice recordings were extracted from semi-standardized video recordings, which captured participants describing what is great about themselves within one minute based on eight life domains (e.g. 'friends', 'sports', 'humour') from which they were asked to choose three (for details see Kordsmeyer & Penke, 2019). The recordings were cut to a length of 5 s, beginning 5 s after the men started to speak. Female participants were recorded reading aloud a standardized voice passage ('rainbow passage', Fairbanks, 1960). The final sentence was used for voice ratings (mean length of the voice stimuli = 7.36 s, $SD = 0.91$ s). Since women's voices were recorded four times (twice per fertility status, see Jünger et al., 2018 and Stern et al., 2019), the recordings concurrent with the face scans from the first infertile session were used. After excluding five recordings for men owing to technical issues and participants not wanting their recordings to be used further, the final stimulus sets comprised 160 recordings for men and 157 for women.

Voice ratings

One-hundred and nineteen raters ($n = 59$ men, age: $M = 20.21$ years, $SD = 4.05$) from a US-American subsample of a larger study on voice perceptions rated men's voice recordings on short- and long-term attractiveness, amongst others (Schild et al., 2020). Raters were equipped with Sennheiser HD 280 Professional headphones and seated in private workstations. Raters provided information on their age, gender, sexual orientation and relationship status as well as on their German language knowledge, which indicated that most raters had no comprehension of German language (95.8% were below the mid-point on a seven-point scale), ensuring that our voice ratings were unbiased by spoken content (but see Baus et al., 2019). The 160 voice recordings were each rated at least 15 times by each sex (only long-term attractiveness was rated by 14 males and 15 females owing to drop out of one male rater) on seven-point Likert scales using the item 'How attractive does the speaker sound for a short-term, uncommitted sexual relationship?' for short-term and 'How attractive does the speaker
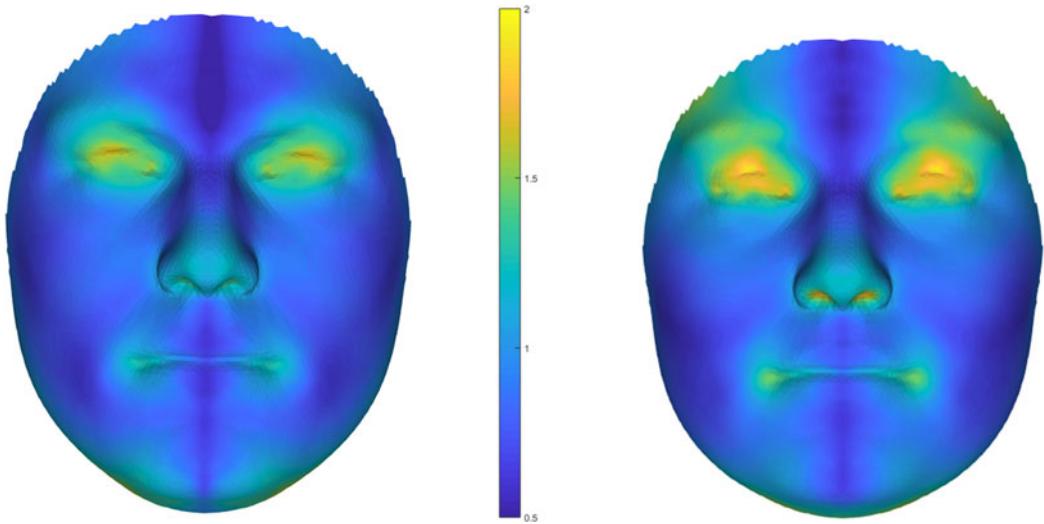
Figure 1. Heat maps showing average facial asymmetries based on 3D face scans for men (left) and women (right). *Note*: $N = 137$ men, $N = 153$ women; unit of the scale is millimetres.

sound for a long-term, committed romantic relationship?' for long-term attractiveness, with the end-points -3 = 'very unattractive' to 3 = 'very attractive'.

Women's voice recordings were judged for short- and long-term attractiveness by overall $N = 42$ German raters ($n = 21$ women, age: $M = 25.00$ years, $SD = 5.04$), equipped with headphones and seated in private workstations. Raters provided information on their age, gender, sexual orientation, relationship status, and whether they were enrolled as a student. The stimulus set ($N = 157$) was split in half to reduce strain on raters, and each set was rated by at least 10 males and females (set 1/2: $n = 80/77$ voice recordings). The items 'How sexually attractive is this woman?' for short-term and 'How attractive is this woman for a long-term romantic relationship?' for long-term attractiveness were assessed on 11-point Likert scales, with the endpoints -5 = 'extremely unattractive' to $+5$ = 'extremely attractive'. Men's and women's voice recordings were presented in randomized order.

Mean ratings (of male and female raters) were used for each item for target men and women (inter-rater reliabilities: intraclass correlations ($ICCs$) > 0.77, all $p < 0.001$). Since ratings of long-term and short-term attractiveness correlated significantly for men ($r = 0.82$, $p < 0.001$) and women ($r = 0.84$, $p < 0.001$), the means of short- and long-term ratings were used for further analyses. On average, women ($M = 0.08$, $SD = 0.93$) received higher ratings of vocal attractiveness than men ($M = -0.28$, $SD = 0.61$; independent-samples t-test: $t = 4.10$, $p < 0.001$, Cohen's $d = 0.46$, for this independent-samples t-test, women's ratings were converted from an eleven-point to a seven-point scale (with the endpoints -3 to $+3$) to ensure comparability with men's ratings).

Face ratings

Standardized facial photographs were taken, while participants stood in front of a white wall directly facing the camera (Canon EOS 350D) at a distance of 2 m and were asked to show a neutral facial expression (with glasses removed). Two photos were taken of each man and overall four of each woman (one for each of the four sessions, see above) and the most suitable photo (regarding neutral expression and head angle) was chosen for the subsequent rating study. Facial photos of 164 men (one photo was excluded owing to issues with the photograph) were presented on computer screens in randomized order. Twelve women (age: $M = 25.2$ years, $SD = 7.1$) rated facial short-term attractiveness (item: 'How sexually attractive is this man?') and long-term attractiveness ('How attractive is this

man for a long-term relationship?') on eleven-point Likert scales with the endpoints −5 = 'extremely unattractive' to +5 = 'extremely attractive'. The raters first previewed the whole sample, with each photograph displayed for 0.5 s, to provide the raters with a first impression of the whole sample. Additionally, for each target man raters were asked to what degree they knew him on a three-point scale (1 = 'not at all', 2 = 'know him by sight', 3 = 'well'). Data points where a rater indicated knowing a given target man well (= 3) were excluded from subsequent analyses. Interrater agreements were high (for short-/long-term attractiveness, $ICCs = 0.86/0.87$). Since ratings of long- and short-term attractiveness were highly correlated ($r = 0.96$, $p < 0.001$), they were aggregated to get a measure comparable with women's global facial attractiveness (see below).

Facial photos of 155 women (two photos were excluded because the subjects did not want their photographs to be used further) were presented on computer screens in randomized order and judged by overall 41 raters (21 women, age: $M = 19.9$ years, $SD = 1.5$) at the University of California, Santa Barbara on global facial attractiveness (one item: 'How attractive is this woman's face?') on an eleven-point Likert scale with the endpoints −5 = 'extremely unattractive' to +5 = 'extremely attractive'. To reduce strain on the raters, a 3 min break was taken after rating half of the stimulus samples. Interrater agreements were very high (for short- and long-term as well as global facial attractiveness and for male and female raters, $ICCs > 0.86$).

Statistical analyses

Bivariate Pearson correlations were calculated for associations between facial FA and vocal attractiveness, separately for men and women. For robustness checks, we ran additional linear regression models, with facial FA as the independent and vocal attractiveness as the dependent variable, controlling for facial attractiveness in one model (as in Hill et al., 2017) and age as well as body mass index (BMI) in another model (as both may be related to or confound FA measures; Manning, 1995; Penke et al., 2009). To examine whether the effect sizes in our sample on associations between facial FA and vocal attractiveness were statistically different from previously reported mean effect sizes, we conducted equivalence tests using the R package *TOSTER* (Lakens, 2017; Lakens et al., 2018). As the smallest effect sizes of interest (equivalence bound) we used the meta-analytic mean effect sizes of $r = −0.20$ for men and $r = −0.26$ for women reported in Hill et al.'s study (2017, only including studies on facial but not bodily FA; see Kordsmeyer & Penke, 2017 for a similar procedure). Additionally, for a more conservative approach we used the confidence intervals' upper limits of the reported mean effect sizes (−0.06 for men and −0.09 for women, Hill et al., 2017; cf. Lakens et al., 2018; Perugini et al., 2014). Hence, we used −0.20 and −0.26 as the regular and −0.06 and −0.09 as the conservative bounds for men and women, respectively, to examine if our study's effect sizes significantly differ from these previously reported results (Lakens et al., 2018).

Results

Descriptive statistics including all ratings of vocal attractiveness and facial FA measures are shown in Table 1, and bivariate correlations between all main variables in Table 2.

Main results

Facial FA and vocal attractiveness were not significantly correlated in men ($r = 0.13$, $p = 0.15$) or women ($r = 0.06$, $p = 0.48$, Table 2, Figure 2). Similarly, when analysing short- and long-term vocal attractiveness separately, associations were non-significant (short-/long-term attractiveness, men: $r = 0.16/0.08$, $p = 0.07/0.36$, women: $r = 0.07/0.04$, $p = 0.41/0.59$). Equivalence tests revealed that, for men, the confidence interval of effect sizes did not include the negative bound of $r = −0.20$ (based on the meta-analytic mean effect size for men reported in Hill et al., 2017), showing that our effect size was statistically different from the previously reported mean effect size ($p < 0.001$, Figure S1). Even when following the more conservative approach and using the upper limit of the

Table 1. Descriptive statistics for all variables measured

Variable	Men			Women		
	M	*SD*	*ICC*	*M*	*SD*	*ICC*
Facial FA	0.98	0.25		0.96	0.31	
Vocal attractiveness	−0.28	0.61	0.77/0.83	0.09	0.93	0.83/0.83
Age	24.3	3.2		23.2	3.4	
BMI	23.98	3.81		22.45	3.34	
Facial attractiveness	−1.35	1.43	0.86/0.87	−0.15	0.99	0.92

Note. ICC, intra-class correlation (interrater reliability); BMI, body mass index; *M*, mean; *SD* = standard deviation. Vocal attractiveness rating scales for men and women: −3 to +3 (women's vocal attractiveness ratings were converted from an 11-point to a seven-point scale (with the endpoints −3 to +3) to ensure comparability with men's ratings for these descriptive statistics); facial attractiveness rating scales for men and women: −5 to +5; ICCs for vocal and facial short-/long-term attractiveness, values of *M* and *SD* for aggregates of short- and long-term attractiveness. *N* = 134–165.

**Table 2.** Bivariate Pearson correlations between all measured variables

Bivariate correlations	Facial FA	Vocal attractiveness	Age	BMI	Facial attractiveness
Facial FA	—	0.06	0.05	−0.02	−0.10
Vocal attractiveness	0.13	—	−0.12	−0.08	0.09
Age	0.14	−0.07	—	0.13	−0.19*
BMI	0.05	−0.08	0.08	—	−0.36***
Facial attractiveness	0.06	0.28***	−0.24**	−0.21**	—

Note. Men (*N* = 134-165) below and women (*N* = 151–157) above the diagonal; *p* < 0.05, **p* < 0.01, ***p* < 0.001.

effect's confidence interval (−0.06 in Hill et al., 2017; cf. Lakens et al., 2018; Perugini et al., 2014), our study's effect size was statistically different (*p* = 0.02, Figure S2). For women, the confidence interval of the effect sizes did not include the negative bound of *r* = −0.26 (Hill et al., 2017), again showing that our effect size was statistically different (*p* < 0.001, Figure S3). Also following the more conservative approach and using the effect's confidence interval's upper limit (−0.09 in Hill et al., 2017), our study's effect size was statistically different (*p* = 0.03, Figure S4). Thus, we provide evidence for a potential absence of an inverse association between facial FA and vocal attractiveness, contrary to previous studies (Abend et al., 2015; Hill et al., 2017).

Robustness analyses

Adding target men's and women's facial attractiveness in one model and age as well as BMI in another model as covariates to linear regression models predicting vocal attractiveness from facial FA (separate models for men and women) revealed only a significant effect for facial attractiveness on vocal attractiveness for men (β = 0.27, *p* < 0.01; other effects, for men: unsigned values of β < 0.16, *p* > 0.09, for women: unsigned values of β < 0.13, *p* > 0.11, Table S4). For target men, three raters self-identified as bisexual or homosexual (for target women, all raters indicated to be heterosexual). Excluding these three raters, the correlation between facial FA and vocal attractiveness remained virtually unchanged (*r* = 0.13, *p* = 0.14). Thus, associations between vocal attractiveness and facial FA were robustly non-significant, also when controlling for raters' sexual orientation, target participants' facial attractiveness or age and BMI.
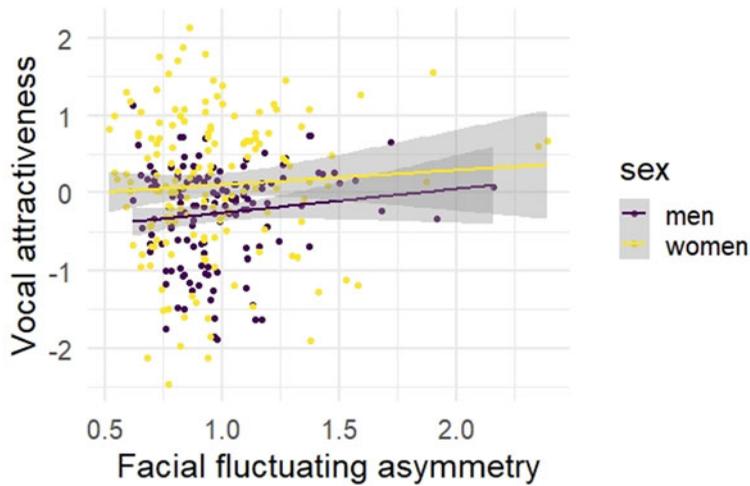
Figure 2. Scatterplot for the associations between facial fluctuating asymmetry and vocal attractiveness, separately for men and women. *Note*: $N = 130$ men, $N = 148$ women; facial fluctuating asymmetry based on automatic measures.

Facial FA from manual landmarks

Employing facial FA measures based on landmarks (vertical, horizontal, and comprehensive facial FA measures, see Table S5 for an overview of facial FA measures) manually placed on 2D photographs (for women only) and 3D scans (for men and women) revealed no significant associations with vocal attractiveness for men (unsigned rs < 0.09, ps > 0.33, Table S1) or women (unsigned rs < 0.14, ps > 0.16, Table S1). For women, manual facial FA measures based on 2D photographs and 3D scans were not significantly correlated (grouped into vertical, horizontal and comprehensive facial FA measures, values of r < 0.19, p > 0.053, Table S2). Some of these manual facial FA measures showed small to medium-sized positive associations with the automatic facial FA measures for women (overall range for 2D/3D: values of r between −0.01/0.19 and 0.21/0.33, values of p between 0.02 and 0.91 for 2D, values of p < 0.02 for 3D, Table S3), but not men (3D: values of r between −0.09 and 0.12, values of p > 0.16, Table S3).

Discussion

Employing accurate and automatic measures of facial FA based on 3D face scans, we found no significant associations with target men's or women's observer-judged vocal attractiveness. These null associations held when controlling for their facial attractiveness as well as age and BMI. Equivalence tests revealed that our study's effect sizes were significantly different from previously reported effects (Hill et al., 2017). Thus, we provide robust evidence against a meaningful inverse relationship of vocal attractiveness with facial FA as a proxy measure of DI (contrary to Abend et al., 2015; Hill et al., 2017; see Kordsmeyer & Penke, 2017 for a similar null finding on bodily FA and mating success).

These results are in stark contrast to earlier studies on facial FA and vocal attractiveness (Abend et al., 2015; Hill et al., 2017), which may partly be explained by differences in methodology. In Hill et al.'s (2017) three studies, facial FA measures were based on relatively few manual landmarks mostly placed on 2D photographs (3D scans were used only for a subsample in study 3), so that their results appear as less reliable and potentially less valid (owing to the few landmarks capturing less shape information compared with automatic measures based on geometric morphometrics; Ekrami et al., 2018). While Abend et al. (2015) employed a geometric morphometrics method for facial FA (which was still landmark-based, however), they only used 2D photographs, thus discarding one

dimension of shape information. Our study's methodology of assessing facial FA is advanced, using automatic geometric morphometric measurements based on 3D face scans (for a validation study see Ekrami et al., 2018), which should deliver more accurate facial FA data, while avoiding human error in manual landmarking (Berssenbrügge et al., 2014; Claes et al., 2011; Ekrami et al., 2018). Moreover, the sample in Abend et al. (2015) was relatively small ($N = 42$ target participants), so that the significant effect could be a false positive owing to the estimation not being very robust (Schönbrodt & Perugini, 2013 recommend sample sizes of at least 250 for stable estimates). Overall, our results are at least as robust and accurate as earlier findings, given our relatively large samples and advanced method of measuring facial FA. Equivalence tests showed that our results were significantly different from these meta-analytic effect sizes for both men and women (Hill et al., 2017), even when more conservatively comparing our effect with the meta-analytic confidence intervals' upper bounds (Lakens et al., 2018). Thus, robust evidence against a meaningful association between facial FA and vocal attractiveness is provided.

Earlier studies suggested a potential confound of associations with facial FA by facial attractiveness. Facial FA may directly affect facial attractiveness, independent of or beyond effects of underlying DI (e.g. Van Dongen & Gangestad, 2011). In the case of vocal attractiveness this confound could only be indirect. For example, presuming an inverse association between facial FA and facial attractiveness (e.g. Grammer & Thornhill, 1994; Komori et al., 2009; Scheib et al., 1999; Simmons et al., 2004), more facially symmetrical people might feel more confident and speak more attractively (Hill et al., 2017). Our study's non-significant relationships between facial FA and vocal attractiveness were robust to controlling for facial attractiveness. Moreover, contrary to earlier evidence (e.g. Grammer & Thornhill, 1994; Komori et al., 2009; Mogilski & Welling, 2017; Rhodes et al., 1998; Scheib et al., 1999), in our study the correlations between facial FA and facial attractiveness were non-significant and close to zero (in the expected negative direction for women, positive for men, see Hughes & Aung, 2018 for suggestions that associations between facial FA and attractiveness may be stronger when assessing attractiveness perceptions based on moving instead of static facial stimuli, see also below). Our results may be seen as more robust than extant findings, owing to a considerably larger sample size (compared with Grammer & Thornhill, 1994; Komori et al., 2009; Scheib et al., 1999) and arguably more accurate assessment of facial FA (automatic 3D-based vs. manual 2D-based measurement). This underlines suggestions that such a link may not be as straightforward and linear as previously assumed and requires further investigation (cf. Jones & Jaeger, 2019; Holzleitner et al., 2019; Simmons et al., 2004).

Our results may speak against vocal attractiveness directly signalling genetic quality in men or women, at least via a link with facial FA. However, vocal characteristics may still convey information about DI and heritable fitness benefits (Hill & Puts, 2016) and play a role in sexual selection (Collins, 2000). First, vocal attractiveness may more strongly relate to measures of bodily FA than facial FA (for an inverse relationship between bodily FA and vocal attractiveness see Hughes et al., 2002 and Abend et al., 2015). Hence, it is recommended to attempt to replicate these findings to establish whether vocal attractiveness may convey information about DI via its link with bodily FA instead of facial FA. Secondly, vocal attractiveness may still indirectly signal genetic quality, considering that it has been debated whether FA provides a direct and reliable measure of DI or not (e.g. Van Dongen & Gangestad, 2011; in Van Dongen et al., 2009 a significant inverse correlation between bodily FA and mating success faded after partialling out DA, the authors concluding that DI does not underlie the effect of bodily FA). As a third alternative, vocal characteristics may play a more direct role in sexual selection, not mediated by asymmetry, by influencing individuals' mating and reproductive success. Previous studies have shown positive effects on reported age of first sexual intercourse, number of sexual partners and extra-pair copulation partners, for example (Hughes et al., 2004; but see Hill et al., 2013 and Kordsmeyer et al., 2018 for null associations of vocal attractiveness with mating success).

Our null findings are strengthened by the fact that they converged for facial FA measured using manual landmarks (based on both 2D photographs and 3D face scans) as well as the automatic and spatially dense approach for both men and women. These manual measurements of facial FA

based on 2D facial photographs and 3D face scans were not significantly correlated (in women), however, replicating Hill et al.'s (2017) null correlation. Moreover, correlations of automatic, spatially dense 3D facial FA measures with manual measures based on 3D scans were descriptively stronger compared with those based on 2D photographs. Given that facial FA measures from 3D scans are described as superior (e.g. Claes et al., 2011; Ekrami et al., 2018), this pattern of results further questions the reliability and validity of earlier findings based on 2D photographs only.

A further concern of this study and similar earlier studies is the validity of facial FA assessments in general. Objectively measuring facial asymmetry from static stimuli (whether 3D scans or 2D photos) may not be particularly ecologically valid, because real-life perceptions of facial characteristics, including asymmetry, may largely be influenced by facial movements (e.g. Rubenstein, 2005). For example, a study showed a non-significant correlation between judgements of men's facial attractiveness based on static vs moving stimuli (Lander, 2008). Another recent study suggests that associations between facial FA and facial attractiveness may increase when attractiveness perceptions are assessed based on moving instead of static facial stimuli (Hughes & Aung, 2018). It was found that, when moving facial stimuli created an appearance of lower facial FA, the target person was rated as being more attractive than when using static photos. Similarly, when moving facial stimuli conveyed greater facial FA, the target person was rated as less attractive compared with attractiveness perceptions based on still photos. Considering only static stimuli as in our study, 3D scans may still be superior to 2D photos as they account for asymmetry information from a third dimension which more closely corresponds to real-life perceptions of moving stimuli. Thus, while assessments of FA based on (static) 3D stimuli provide at least somewhat increased ecological validity, future studies should additionally assess asymmetry (and attractiveness) perceptions based on moving stimuli in order to achieve even more ecologically valid results.

In addition to these confounds pertaining to facial FA assessments, our findings may be limited by the fact that we employed standardized voice recordings for women only (rainbow passage, Fairbanks, 1960; this was used by Hill et al., 2017 for their male target participants as well). For men, only partly standardized voice recordings were available, contrary to some previous studies (e.g. Abend et al., 2015; Hill et al., 2017). For our voice recordings, a range of topics was provided to choose from (e.g. 'friends', 'sports', 'family'), but the men were free to talk about what they liked for up to 1 min (see Kordsmeyer & Penke, 2019 for more details and a further discussion of these recordings). This means that whereas for women the content and prosody of the voice recordings were controlled, for men they were not. Previous research has shown that the type of stimulus (verb vs. vowel) influences observer perceptions, such as for ratings of vocal attractiveness (e.g. Ferdenzi et al., 2013; but see Mahrholz et al.,, 2018 and Puts et al., 2012 reporting great convergence of vocal parameters across different voice recording). In addition, in our study the native languages of the target participants and raters differed for men, which should have reduced the bias arising from differences in spoken content. However, an earlier study showed that personality perceptions were very similar whether the listeners' and target participants' native languages matched or not (Baus et al., 2019). This means that the effect of most raters not being able to understand the content of the only partly standardized voice recordings may not be as strong as initially proclaimed. Moreover, our items for ratings of vocal and facial attractiveness were not entirely consistent between men and women. Firstly, slightly different wordings for items assessing short- and long-term vocal attractiveness were used. Secondly, for facial attractiveness, one item assessing global attractiveness was employed for women, whereas for men two items on attractiveness for a short- and long-term relationship were used. This means that for vocal and facial attractiveness slightly different constructs may have been measured for men and women, so that results are not completely comparable. While unstandardized voice recordings are not necessarily less valid for the assessment of vocal attractiveness, it is a limitation of the current study that they are less comparable with the female sample and previous studies. Initial evidence for the validity of our study's men's vocal attractiveness ratings is provided by a significant medium-sized correlation with observer-rated facial attractiveness in line with the one-ornament hypothesis (for women we found no such positive association contrary to some earlier findings, Feinberg et al., 2005; for a review see Feinberg, 2008; for conflicting evidence questioning the one ornament hypothesis

see Lander, 2008; Rezlescu et al., 2015; Zuckerman & Sinicropi, 2011). Moreover, ratings of men's vocal attractiveness appeared to be reliable (in terms of high interrater agreements, comparable with those for women). Still, future studies should strive to replicate our findings using consistent items and standardized voice recordings for both men and women.

To conclude, contrary to earlier studies (Abend et al., 2015; Hill et al., 2017), we provide robust evidence against a meaningful association between men's and women's facial fluctuating asymmetry and vocal attractiveness, as underlined by equivalence tests. Facial fluctuating asymmetry was measured using an arguably superior, automatic spatially dense approach based on 3D face scans. Additionally, we did not detect significant correlations using a variety of more conservative measures (manual landmarks based on 2D photographs and 3D scans) or when controlling for facial attractiveness, age and BMI. Thus, it seems that individuals' vocal attractiveness does not convey information on their heritable fitness benefits, at least not as manifested by facial fluctuating asymmetry as a proxy measure of developmental instability.

References

Abend, P., Pflüger, L. S., Koppensteiner, M., Coquerelle, M., & Grammer, K. (2015). The sound of female shape: A redundant signal of vocal and facial attractiveness. *Evolution and Human Behavior*, 36, 174–181. https://doi.org/10.1016/j.evolhumbehav.2014.10.004

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, 9. https://doi.org/10.1038/s41598-018-36518-6

Berssenbrügge, P., Berlin, N. F., Kebeck, G., Runte, C., Jung, S., Kleinheinz, J., & Dirksen, D. (2014). 2D and 3D analysis methods of facial asymmetry in comparison. *Journal of Cranio-Maxillofacial Surgery*, 42, e327–e334. https://doi.org/10.1016/j.jcms.2014.01.028

Brown, W. M., Price, M. E., Kang, J., Pound, N., Zhao, Y., & Yu, H. (2008). Fluctuating asymmetry and preferences for sex-typical bodily characteristics. *Proceedings of the National Academy of Sciences*, 105, 12938–12943. https://doi.org/10.1073/pnas.0710420105

Claes, P., Walters, M., Vandermeulen, D., & Clement, J. G. (2011). Spatially-dense 3D facial asymmetry assessment in both typical and disordered growth. *Journal of Anatomy*, 219, 444–455. https://doi.org/10.1111/j.1469-7580.2011.01411.x

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, 60, 773–780. https://doi.org/10.1006/anbe.2000.1523

Ekrami, O., Claes, P., White, J. D., Zaidi, A. A., Shriver, M. D., & Van Dongen, S. (2018). Measuring asymmetry from high-density 3D surface scans: An application to human faces. *PLoS ONE*, 13, e0207895. https://doi.org/10.1371/journal.pone.0207895

Fairbanks, G. (1960). *Voice and articulation drillbook*. Harper & Row.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology: Issues, News, and Reviews*, 17, 112–118. https://doi.org/10.1002/evan.20166

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Smith, M. J. L., Cornwell, R. E., … Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26, 398–408. https://doi.org/10.1016/j.evolhumbehav.2005.04.001

Ferdenzi, C., Patel, S.,Mehu-Blantar, I., Khidasheli, M., Sander, D., & Delplanque, S. (2013). Voice attractiveness: Influence of stimulus duration and type. *Behavior Research Methods*, 45, 405–413. https://doi.org10.3758/s13428-012-0275-0

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2005). Women's sexual interests across the ovulatory cycle depend on primary partner developmental instability. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 2023–2027. https://doi.org/10.1098/rspb.2005.3112

Gangestad, S. W., & Thornhill, R. (1997). Human sexual selection and developmental stability. In J. A. Simpson, & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 169–195). Lawrence Erlbaum.

Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings of the Royal Society of London B: Biological Sciences*, 265, 927–933. https://doi.org/10.1098/rspb.1998.0380

Graham, J., & Özener, B. (2016). Fluctuating asymmetry of human populations: A review. *Symmetry*, 8, 154. https://doi.org/10.3390/sym8120154

Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, 108, 233–242. http://dx.doi.org/10.1037/0735-7036.108.3.233

Hill, A. K., Cárdenas, R. A., Wheatley, J. R., Welling, L. L., Burriss, R. P., Claes, P., … Puts, D. A. (2017). Are there vocal cues to human developmental stability? Relationships between facial fluctuating asymmetry and voice attractiveness. *Evolution and Human Behavior*, 38, 249–258. https://doi.org/10.1016/j.evolhumbehav.2016.10.008

Hill, A. K., Hunt, J., Welling, L. L., Cárdenas, R. A., Rotella, M. A., Wheatley, J. R., … Puts, D. A. (2013). Quantifying the strength and form of sexual selection on men's traits. *Evolution and Human Behavior*, 34, 334–341. https://doi.org/10.1016/j.evolhumbehav.2013.05.004

Hill, A. K., & Puts, D. A. (2016). Vocal attractiveness. In V. Weekes-Shackelford, T. K. Shackelford, & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science*. Springer International.

Holzleitner, I. J., Lee, A. J., Hahn, A. C., Kandrik, M., Bovet, J., Renoult, J. P., … Jones, B. C. (2019). Comparing theory-driven and data-driven attractiveness models using images of real women's faces. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 1589–1595. https://psycnet.apa.org/doi/10.1037/xhp0000685

Hughes, S. M., & Aung, T. (2018). Symmetry in motion: Perception of attractiveness changes with facial movement. *Journal of Nonverbal Behavior*, 42, 267–283. https://doi.org/10.1007/s10919-018-0277-4

Hughes, S. M., Dispenza, F., & Gallup Jr, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, 25, 295–304. https://doi.org/10.1016/j.evolhumbehav.2004.06.001

Hughes, S. M., Harrison, M. A., & Gallup Jr, G. G. (2002). The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior*, 23, 173–180. https://doi.org/10.1016/S1090-5138(01)00099-X

Hughes, S. M., Pastizzo, M. J., & Gallup, G. G. (2008). The sound of symmetry revisited: Subjective and objective analyses of voice. *Journal of Nonverbal Behavior*, 32, 93–108. https://doi.org/10.1007/s10919-007-0042-6

Jones, A. L., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, 11, 279. https://doi.org/10.3390/sym11020279

Jünger, J., Kordsmeyer, T. L., Gerlach, T. M., & Penke, L. (2018). Fertile women evaluate male bodies as more attractive, regardless of masculinity. *Evolution and Human Behavior*, 39, 412–423. https://doi.org/10.1016/j.evolhumbehav.2018.03.007

Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution*, 56, 1909–1920. https://doi.org/10.1111/j.0014-3820.2002.tb00117.x

Komori, M., Kawamura, S., & Ishihara, S. (2009). Averageness or symmetry: Which is more important for facial attractiveness? *Acta Psychologica*, 131, 136–142. https://doi.org/10.1016/j.actpsy.2009.03.008

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra- and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior*, 39, 424–436. https://doi.org/10.1016/j.evolhumbehav.2018.03.008

Kordsmeyer, T., & Penke, L. (2019). Effects of male testosterone and its interaction with cortisol on self- and observer-rated personality states in a competitive mating context. *Journal of Research in Personality*, 78, 76–92. https://doi.org/10.1016/j.jrp.2018.11.001

Kordsmeyer, T. L., & Penke, L. (2017). The association of three indicators of developmental instability with mating success in humans. *Evolution and Human Behavior*, 38, 704–713. https://doi.org/10.1016/j.evolhumbehav.2017.08.002

Lakens, D. (2017). TOSTER: Two one-sided tests (TOST) equivalence testing (version 0.3.4) [Computer software]. Retrieved from https://CRAN.R-project.org/package=TOSTER

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1, 259–269. https://doi.org/10.1177/2515245918770963

Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, 75, 817–822. https://doi.org/10.1016/j.anbehav.2007.07.001

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PloS One*, 13. https://doi.org/10.1371/journal.pone.0204991

Manning, J. T. (1995). Fluctuating asymmetry and body weight in men and women: Implications for sexual selection. *Ethology and Sociobiology*, 16, 145–153. https://doi.org/10.1016/0162-3095(94)00074-H

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*, 36, 235–247. https://doi.org/10.1007/s11692-009-9055-x

Mogilski, J. K., & Welling, L. L. (2017). The relative importance of sexual dimorphism, fluctuating asymmetry, and color cues to health during evaluation of potential partners' facial photographs. *Human Nature*, 28, 53–75. https://doi.org/10.1007/s12110-016-9277-4

Palmer, A. R., & Strobeck, C. (2003). Fluctuating asymmetry analyses revisited. In M. Polak (Ed.), *Developmental instability: Causes and consequences* (pp. 279–319). Oxford University Press.

Penke, L., Bates, T. C., Gow, A. J., Pattie, A., Starr, J. M., Jones, B. C., … Deary, I. J. (2009). Symmetric faces are a sign of successful cognitive aging. *Evolution and Human Behavior*, 30, 429–437. https://doi.org/10.1016/j.evolhumbehav.2009.06.001

Pertoldi, C., Podesta, M., Loeschcke, V., Schandorff, S., Marsili, L., Manusci, C., … Randi, E. (2000). Effect of the 1990 die-off in the northern Italian seas on the developmental stability of the striped dolphin *Stenella coeruleoalba* (Meyen, 1833). Biological *Journal of the Linnean Society*, 71, 61–70. https://doi.org/10.1111/j.1095-8312.2000.tb01242.x

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, 9, 319–332. https://doi.org/10.1177/1745691614528519

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, 279, 601–609. https://doi.org/10.1098/rspb.2011.0829

Puts, D. (2016). Human sexual selection. *Current Opinion in Psychology*, 7, 28–32. https://doi.org/10.1016/j.copsyc.2015.07.011

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, 49, 227–243. https://doi.org/10.1080/00224499.2012.658924

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, 39, 355–370. https://doi.org/10.1007/s10919-015-0214-8

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, 5, 659–669. https://doi.org/10.3758/BF03208842

Roldan, E. R. S., Cassinello, J., Abaigar, T., & Gomendio, M. (1998). Inbreeding, fluctuating asymmetry, and ejaculate quality in an endangered ungulate. *Proceedings of the Royal Society of London B: Biological Sciences*, 265, 243–248. https://doi.org/10.1098/rspb.1998.0288

Rubenstein, A. J. (2005). Variation in perceived attractiveness: Differences between dynamic and static faces. *Psychological Science*, 16, 759–762. https://doi.org/10.1111/j.1467-9280.2005.01610.x

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London B: Biological Sciences*, 266, 1913–1917. https://doi.org/10.1098/rspb.1999.0866

Schild, C., Aung, T., Kordsmeyer, T. L., Cárdenas, R. A., Puts, D. A., & Penke, L. (2020). *Linking human male vocal parameters to perceptions, body morphology, strength and hormonal profiles in contexts of sexual selection.*

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47, 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Simmons, L. W., Rhodes, G., Peters, M., & Koehler, N. (2004). Are human preferences for facial symmetry focused on signals of developmental instability? *Behavioral Ecology*, 15, 864–871. https://doi.org/10.1093/beheco/arh099

Singh, D., & Rosen, V. C. (2001). Effects of maternal body morphology, morning sickness, gestational diabetes and hypertension on fluctuating asymmetry in young women. *Evolution and Human Behavior*, 22, 373–384. https://doi.org/10.1016/S1090-5138(01)00082-4

Stern, J., Schild, C., Arslan, R. C., Jones, B. C., DeBruine, L. M., Hahn, A., … Penke, L. (2019, April 11). Do voices carry valid information about a speaker's personality? https://doi.org/10.31234/osf.io/9a68c

Thornhill, R., & Gangestad, S. W. (1994). Human fluctuating asymmetry and sexual behavior. *Psychological Science*, 5, 297–302. https://doi.org/10.1111/j.1467-9280.1994.tb00629.x

Van Dongen, S., Cornille, R., & Lens, L. (2009). Sex and asymmetry in humans: What is the role of developmental instability? *Journal of Evolutionary Biology*, 22, 612–622. https://doi.org/10.1111/j.1420-9101.2008.01667.x

Van Dongen, S., & Gangestad, S. W. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior*, 32, 380–398. https://doi.org/10.1016/j.evolhumbehav.2011.03.002

Van Valen, L. (1962). A study of fluctuating asymmetry. *Evolution*, 16, 125–142. https://doi.org/10.1111/j.1558-5646.1962.tb03206.x

Yeo, R. A., Gangestad, S. W., Edgar, C., & Thoma, R. (1999). The evolutionary genetic underpinnings of schizophrenia: The developmental instability model. *Schizophrenia Research*, 39, 197–206. https://doi.org/10.1016/S0920-9964(99)00074-2

Zuckerman, M., & Sinicropi, V. (2011). When physical and vocal attractiveness differ: Effects on favorability of interpersonal impressions. *Journal of Nonverbal Behavior*, 35, 75–86. https://doi.org/10.1007/s10919-011-0106-5