



HAL
open science

Voice-based judgments of sex, height, weight, attractiveness, health and psychological traits based on free speech versus scripted speech

Piotr Sorokowski, Katarzyna Pisanski, Tomasz Frąckowiak, Aleksander Kobylarek, Agata Groyecka-Bernard

► To cite this version:

Piotr Sorokowski, Katarzyna Pisanski, Tomasz Frąckowiak, Aleksander Kobylarek, Agata Groyecka-Bernard. Voice-based judgments of sex, height, weight, attractiveness, health and psychological traits based on free speech versus scripted speech. *Psychonomic Bulletin and Review*, In press. hal-04367406v1

HAL Id: hal-04367406

<https://hal.science/hal-04367406v1>

Submitted on 30 Dec 2023 (v1), last revised 18 Jan 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PSYCHONOMIC BULLETIN & REVIEW

Voice-based judgments of sex, height, weight, attractiveness, health and psychological traits based on free speech versus scripted speech

Journal:	<i>Psychonomic Bulletin & Review</i>
Manuscript ID	PBR-BR-22-320.R1
Manuscript Type:	Brief Report
Date Submitted by the Author:	29-Nov-2023
Complete List of Authors:	Sorokowski, Piotr; University of Wrocław, Institute of Psychology Pisanski, Katarzyna; University of Wrocław, Frackowiak, Tomasz; University of Wrocław, Institute of Psychology Kobylarek, Aleksander; University of Wrocław, Institute of Pedagogy Groyecka-Bernard, Agata; University of Wrocław, Institute of Psychology
Keywords:	voice perception, voice recording, voice-based judgements, speech

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Voice-based judgments of sex, height, weight, attractiveness, health and psychological traits based on free speech versus scripted speech.

Piotr Sorokowski¹, Katarzyna Pisanski^{1,2,3*}, Tomasz Frąckowiak¹, Aleksander Kobylarek⁴,
Agata Groyecka-Bernard¹

¹Institute of Psychology, University of Wrocław, Wrocław, Poland

²ENES Bioacoustics Research Lab, CRNL, CNRS UMR5292, University Jean Monnet, Saint-
Etienne, France.

³Centre National de la Recherche Scientifique, Laboratoire Dynamique du Langage,
Université Lyon 2, France

⁴ Institute of Pedagogy, University of Wrocław, Wrocław, Poland

Author Note

Piotr Sorokowski: <https://orcid.org/0000-0001-9225-9965>

Katarzyna Pisanski: <https://orcid.org/0000-0003-0992-2477>

Tomasz Frąckowiak: <https://orcid.org/0000-0002-0244-2879>

Aleksander Kobylarek: <https://orcid.org/0000-0002-4562-9035>

Agata Groyecka-Bernard: <https://orcid.org/0000-0002-1932-4828>

*Corresponding author: Katarzyna Pisanski kasiapisanski@gmail.com;

Correspondence concerning this article should be addressed to Katarzyna Pisanski, Institute of
Psychology, University of Wrocław, ul. Dawida 1, 50-527 Wrocław, Poland.

Funding: This work was supported by a grant from the National Science Centre (grant no.
2016/23/B/HS6/00771).

Abstract

How do we perceive others based on their voices? This question has attracted research and media attention for decades, producing hundreds of studies showing that the voice is socially and biologically relevant, but these studies vary in methodology and ecological validity. Here we test whether vocalizers producing **read** versus **free** speech are judged similarly by listeners on ten biological and/or **psychosocial** traits. In perception experiments using speech from 208 men and women and ratings from 4088 listeners, we show that listeners' assessments of **vocalizer sex and age are highly accurate, regardless of speech type. Assessments of femininity-masculinity and health also do not differ between free and read speech. In contrast, read speech elicits higher ratings of attractiveness and dominance, whereas free speech elicits higher ratings of likeability and trustworthiness. Importantly, these differences were small, and we additionally show moderate to strong correlations between ratings of the same vocalizers based on their read and free speech for all ten traits, indicating that voice-based judgements are highly consistent within speakers, whether or not speech is spontaneous.** Our results provide evidence that the human voice can communicate various biological and psychosocial traits via both **read** and **free** speech, with theoretical and practical implications.

Keywords: voice perception; voice recording; voice-based judgements; speech; communication

Introduction

Vocal cues are a valuable source of socially relevant information about others (Kamiloğlu, & Sauter, 2021; Latinus, & Belin, 2011). Their meaning, evolutionary significance, and social effects have been emphasized across diverse disciplines including biology, psychology, anthropology, ethology, linguistics, the computer sciences and media studies. Given the increasing quantity of studies on voice perception, there is a growing need to better understand how the methodologies applied to investigate human nonverbal vocal communication may influence research outcomes (Groyecka-Bernard et al., 2022; Lavan, 2023; Pisanski et al., 2021; Sorokowski et al., 2022).

One prevalent paradigm in voice research is to use perception/playback experiments to assess listeners' evaluations of unseen speakers based only on their voices. Such studies have, for example, investigated listeners' assessments of social or psychological characteristics of the speakers, such as dominance (Puts et al., 2007), cooperativeness (Knowles & Little, 2016), or authority (Sorokowski et al., 2019), and biological traits including body size or physical strength (reviewed in Pisanski & Bryant, 2019; Aung & Puts, 2019). These experiments consistently highlight the social value of vocal cues in human communication, however research into methodological aspects of such studies is scarce despite considerable variation in methodological protocols across studies, and particularly in the types of voice stimuli used to measure listeners' voice-based assessments.

Although we know that even extremely short speech stimuli, such as a single word, can encode a great deal of socially relevant information about the speaker (McAleer et al., 2014), few researchers have systematically tested how the perception or accuracy of voice ratings is influenced by variations in the length and type of voice samples (e.g., Ferdenzi et al., 2013; Mahrholz et al., 2018, Painter et al., 2021; Groyecka-Bernard et al., 2022). In one of the earliest studies to test this, Ferdenzi and colleagues (2013) showed that voice

1
2
3 69 attractiveness judgments were higher for words than for vowel sounds. Mahrholz and
4
5 70 colleagues (2018) later showed that listeners' voice-based ratings of trustworthiness,
6
7 71 dominance, and attractiveness were highly correlated when based on a word versus sentence
8
9 72 produced by the same speakers, and did not depend on whether speech was socially relevant
10
11 73 or neutral. Painter and colleagues (2021) found that accuracy in listeners' perceptions of
12
13 74 sexual orientation from speech, though generally low, varied depending on whether speakers
14
15 75 were producing a single scripted word or one or two scripted sentences. Most recently,
16
17 76 Groyecka-Bernard and colleagues (2022) showed that listeners' perceptions of several
18
19 77 socially relevant traits including attractiveness, trustworthiness, dominance, likability, health,
20
21 78 masculinity and femininity increased as the length of speech stimuli increased. Listeners'
22
23 79 ratings of these traits were slightly higher for sentence-long greetings or read renditions of the
24
25 80 first paragraph of the classic, content-neutral Rainbow Passage, than they were for vowels and
26
27 81 single words. Nevertheless, listeners' judgments were highly similar for the same speakers,
28
29 82 regardless of the length of their scripted speech, indicating that socially relevant perceptions
30
31 83 of speakers are not wholly altered but instead are moderated by the length of their speech.
32
33 84 Critically, all of these studies focused on scripted or read speech stimuli, with none having
34
35 85 included free spontaneous speech, despite known differences in the way that people talk when
36
37 86 they read versus speak freely (e.g., Szekely et al., 2020; Winkworth, 1994).

37
38
39
40
41
42
43
44 87 It thus remains unknown whether freely produced spontaneous speech elicits different
45
46 88 impressions of speakers than does standardized scripted or read speech, a key aim of the
47
48 89 present study. This is an important research question both theoretically and methodologically.
49
50 90 Theoretically, answering this question can provide insight into the extent to which acoustic
51
52 91 cues to speaker traits are encoded in different kinds of speech, as we predict they will be,
53
54 92 given that socially-rich acoustic parameters such as voice pitch are preserved across speech
55
56 93 types and even nonverbal vocalizations within speakers (Pisanski et al., 2020, 2021).

1
2
3 94 Methodologically, answering this question can clarify whether studies using different kinds of
4
5 95 speech stimuli are comparable, and indeed, whether using free versus read speech is preferred
6
7 96 in terms of ensuring the more robust responses from listeners. Practically, this research can
8
9 97 also uncover whether reading text or speeches, which many public speakers in politics and
10
11 98 broadcasting do, causes listeners to perceive us more or less favourably.
12
13

14 99 The large body of existing research on human voice production and perception is
15
16 100 highly inconsistent in terms of methodology and voice stimulus choice. Some research
17
18 101 designs provide a high degree of standardization by ensuring that all participants produce
19
20 102 vocal stimuli with the same verbal content (e.g., Atkinson et al., 2012, Knowles & Little,
21
22 103 2016; Pipitone & Gallup, 2008; and many others), while others prioritize ecological validity
23
24 104 over internal validity and therefore record spontaneous, unscripted free-speech sentences (e.g.
25
26 105 Schild et al., 2020; Puts et al., 2007; Tigue, et al., 2012, Mayew et al., 2013, and many
27
28 106 others). Ecological validity refers to the extent to which the findings of a research study can
29
30 107 be generalized to real-life settings, and it is generally clear that listening to a free speech
31
32 108 sample more closely reflects a real-life social interaction than does listening to a standardized
33
34 109 recording of someone reading, particularly if the read text is socially irrelevant and affectively
35
36 110 neutral, as it most often is.
37
38
39
40

41
42 111 At the same time, it is also true that the free speech design bears certain risks, as there
43
44 112 are many factors that can influence the nonverbal vocal parameters or prosody of speech and
45
46 113 thus perceptions of the freely speaking person. Indeed, according to previous studies, free
47
48 114 speech has certain characteristics that distinguish it from read speech (Szekely et al., 2020;
49
50 115 Winkworth, 1994). For example, free speech tends to have a higher proportion of filled
51
52 116 pauses such as 'um' and 'uh' (Szekely et al., 2019), as well as more variation in rhythm,
53
54 117 intonation, speaking rate and pitch (Wester, et al., 2016). The relatively greater variance in
55
56 118 free speech might emphasize psycho-biological traits of speakers, namely by drawing
57
58
59
60

1
2
3 119 attention to stable individual differences in vocal parameters that can index speaker traits,
4
5 120 such as dominance or age, which are often indexed by individual differences in mean voice
6
7 121 pitch (Kreiman & Sidtis, 2011). This could in turn increase accuracy in listeners' assessments
8
9 122 of such traits based on free speech compared to read speech. In contrast, increased vocal
10
11 123 variance might occlude or distract from otherwise reliable vocal indices of speaker traits, for
12
13 124 example by drawing attention away from relatively stable vocal markers toward more
14
15 125 idiosyncratic speaking patterns that are less biologically constrained and thus less biologically
16
17 126 relevant, such as speaking rate.

18
19
20
21 127 The aim of the current study was to test whether several commonly studied traits,
22
23 128 including social, biological and psychological characteristics of men and women, would be
24
25 129 assessed differently by listeners from the voice depending on whether these same vocalizers
26
27 130 were producing read or free speech with neutral content. We thus compared listeners'
28
29 131 assessment of attractiveness, trustworthiness, dominance, likability, femininity/masculinity
30
31 132 and health as well as accuracy in their judgments of vocalizer sex, age, height, and weight
32
33 133 based on voice recordings of free speech versus standardized read speech produced by the
34
35 134 same men and women.

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55

136 **Methods**

137 The study was conducted in accordance with the Declaration of Helsinki. Study
138 protocols were accepted by the ethics committee at the Institute of Psychology, University of
139 [masked for peer review]. Vocalizers and raters provided informed consent prior to study
140 inclusion. Vocalizers were informed that their voice recordings will be used for research
141 purposes and that they will be played to other participants in perception experiments.

142 **Participants**

143 *Vocalizers*

56
57
58
59
60

1
2
3 144 Voice samples for the current study were recorded from 208 adult men and women
4
5 145 ($M_{age} = 30.4$, $SD_{age} = 11.54$; 49% women), hereafter referred to as vocalizers. We aimed for a
6
7 146 sample size of approximately 100 vocalizers per sex, based on a trade-off between the sample
8
9 147 size of raters that would be required to judge such a large sample of speech stimuli (totaling
10
11 148 416) while also retaining a representative sample of men and women with variable
12
13 149 interindividual differences in the traits of interest and in nonverbal features of the voice.
14
15 150 Vocalizers were recruited through snowball sampling by researchers, research assistants, their
16
17 151 colleagues and acquaintances. All vocalizers were native Polish speakers and were not
18
19 152 compensated for study participation.
20
21
22

23 153 *Voice raters*

24
25
26 154 A separate group of participants took part in the study as listeners (voice raters), with a
27
28 155 final sample size of 4088 men and women ($M_{age} = 32.16$, $SD_{age} = 12.97$, 46% women), after
29
30 156 removal of approximately 5% of data from participants who failed listening or attention
31
32 157 checks (see below). The majority of raters (67% or $n = 2446$, with 51.9% women, $M_{age} =$
33
34 158 26.2 , $SD_{age} = 10.3$) were recruited by the research team using the snowball sampling method,
35
36 159 wherein researchers posted recruitment ads on their social media profiles and around their city
37
38 160 of residence, both inside and outside of the university, aiming to recruit a representative
39
40 161 sample (i.e., including older individuals and non-students). These participants were
41
42 162 reimbursed through a lottery draw of small prizes such as pen drives, hard drives, and
43
44 163 thermoses. To further increase the representativeness of our rater sample, the remainder of
45
46 164 raters ($n = 1642$, with 37.6% women, $M_{age} = 41.1$, $SD_{age} = 12.2$) were recruited with the help
47
48 165 of the research recruitment firm (Imas International). These participants completed the
49
50 166 experiment online and were reimbursed in cash. The number of raters was determined such
51
52
53
54
55 167 that each voice stimulus would be rated approximately 20 times, as high interrater agreement
56
57
58
59
60

1
2
3 168 (alphas > 0.80, $p < .001$) among judges can typically be achieved with 15 raters (see
4
5 169 Kordsmeyer et al., 2018; Schild et al., 2020).
6
7
8 170

10 171 **Procedure**

12 172 *Voice recording*

14 173 Voice recording took place in a quiet room. Vocalizers were invited to take part in a
16 174 recording session. Prior to voice recording, each vocalizer completed a short survey, and their
18 175 height and weight were measured using metric tape and a scale. Afterwards, they were
20 176 requested to provide specified voice samples (see: voice stimuli) that were recorded using a
22 177 Zoom H4n recorder microphone at a distance of 10 cm from the mouth. Voice recordings
24 178 were saved as WAV files at 96 kHz sampling frequency and 16-bit resolution, and later
26 179 transferred to a laptop. After recording the voice samples, participants were thanked and
28 180 debriefed about the study aims.
30
31

33 181 *Voice stimuli*

35 182 Each vocalizer provided eight randomized voice samples, among which two were
37 183 analyzed for the purposes of the current study, namely:

- 40 184 1. Reading a short, scripted paragraph describing a weather forecast (read speech)
- 42 185 2. Free speech about the same topic, the current weather conditions (free speech)

44 186 *Perception experiment*

46 187 Listeners in the perception experiment rated voice stimuli through a dedicated web
48 188 platform. The recorded voice samples (2 per vocalizer) were incorporated into an online web
50 189 app prepared for the purpose of this study. Raters were invited to complete the survey at the
52 190 University or in their homes. Those who took part in the rating sessions at the University
54 191 completed the survey either individually or in small groups in a designated room (allowing for
56 192 a sufficient distance between the participants that assured their privacy and independence in
58
59
60

1
2
3 193 listeners' responses). Those who completed the survey at home were asked to avoid other
4
5 194 competing noises or distractions while performing the assessments. Lab raters used
6
7 195 Sennheiser HD 210 professional headphones to listen to stimuli, while participants who
8
9 196 completed the playback online were instructed to use high quality headphones.
10
11

12 197 At the beginning of the session, raters were exposed to a test voice sample to ensure
13
14 198 they can hear the stimulus properly and understand the study instructions. Additionally, the
15
16 199 survey included one attention check item placed in the middle of the survey (instead of a
17
18 200 voice sample to rate, listeners heard the instructions: "This is an attention checking question –
19
20 201 please mark 1"). Approximately 5% of participants failed the entry test or attention check and
21
22 202 thus their data were excluded from further analysis.
23
24
25

26 203 For each rater, a sample of 10 vocalizers (5 per sex), producing both read and free
27
28 204 speech, was randomly drawn from the pool of all 208 vocalizers. Thus, each rater heard a total
29
30 205 of 20 speech stimuli. Each speech stimulus was judged independently on a single trial and for
31
32 206 a single vocalizer characteristic and was played in full to listeners before a rating could be
33
34 207 given. Raters were randomly assigned to assess only one of the following ten characteristics
35
36 208 for all voice stimulus samples presented to them: sex/gender (male or female); age in years
37
38 209 (How old is this person?); height in centimeters (How tall is this person?); weight in
39
40 210 kilograms (How much does this person weigh?); attractiveness on a 1 to 7 scale, with 1
41
42 211 indicating very unattractive and 7 indicating very attractive (How attractive is this person?);
43
44 212 dominance on a 1 to 7 scale, with 1 indicating not dominant at all and 7 indicating very
45
46 213 dominant (How dominant is this person?); trustworthiness on a 1 to 7 scale, with 1 indicating
47
48 214 not trustworthy at all and 7 indicating very trustworthy (How trustworthy is this person?);
49
50 215 likability on a 1 to 7 scale, with 1 indicating not likable at all and 7 indicating very likable
51
52 216 (How likable is this person?); femininity/masculinity on a 1 to 7 scale, with 1 indicating very
53
54 217 feminine and 7 indicating very masculine (How feminine/masculine is this person?); and
55
56
57
58
59
60

1
2
3 218 finally, health on a 1 to 7 scale, with 1 indicating very unhealthy and 7 indicating very healthy
4
5 219 (How healthy is this person?). Each speech stimulus was, therefore, evaluated by a mean of
6
7 220 20 raters ($M = 19.57$, $SD = 4.75$ see supplementary materials). Finally, participants completed
8
9 221 a short demographic survey and were thanked and debriefed. The procedure took
10
11 222 approximately 20 to 30 minutes.
12
13
14
15
16

17 224 **Statistical analyses**

18
19 225 To examine the effect of speech stimulus type on perceived traits of the vocalizers, as
20
21 226 well as on the accuracy of listeners' ratings of biological traits, we proceeded with a series of
22
23 227 Linear Mixed Models (using Restricted Maximum Likelihood estimator) with responses
24
25 228 clustered both within speakers and raters. Because of sizeable sexual dimorphism in the
26
27 229 acoustic properties of the human voice (Titze, 1989), we conducted separate models for male
28
29 230 and female vocalizers. Most models included ratings of a given trait as the outcome variable,
30
31 231 speech type (categorical variable coded 0 – read, 1 - free), method of data collection
32
33 232 (categorical variable coded 0 – online, 1 – lab), vocalizer age, rater gender (categorical
34
35 233 variable coded 0-F, 1-M) and rater age as fixed variables, and vocalizers' and raters' coded
36
37 234 identities as random effects. In the model examining the effect of speech type on the accuracy
38
39 235 of sex/gender ratings, we entered accuracy in sex assessment as the outcome variable (0-
40
41 236 incorrect, 1-correct). Additionally, in the models with age, height, and weight as outcomes,
42
43 237 we included the actual height or weight of the speakers (as the relationship between actual and
44
45 238 rated age reflects the accuracy of the ratings) and age/height/weight*speech type as
46
47 239 interaction effects, to test whether the accuracy of listeners' ratings differs between read and
48
49 240 free speech. All continuous predictors were grand-mean centred. All model results are
50
51 241 presented in Tables S1-S11 in the supplementary materials. For means and standard
52
53 242 deviations of actual properties and rated properties (by speech type), see Tables S12 and S13,
54
55
56
57
58
59
60

1
2
3 243 respectively, in the supplementary materials. We also conducted an additional analysis with
4
5 244 an variable, namely the length of the vocal stimulus. Since the results of this analysis closely
6
7
8 245 resemble those presented below (supplementary materials, pages 25-52), we have included
9
10 246 the less extensive findings in the main article.

11
12 247 Due to multiple comparisons, we performed Bonferroni correction and set the alpha
13
14 248 level to 0.005 (0.05/10). Estimates presented in the main text were not standardized. For
15
16
17 249 standardized betas see Figure 1. All analyses and figures were performed using Jamovi, R
18
19 250 using packages ggplot2 (Wickham et al., 2016) and parameters (Lüdecke, 2020) and Python
20
21 251 using matplotlib (Ari, 2014) package. Raw data and codebooks for Jamovi and scripts can be
22
23
24 252 retrieved from: https://osf.io/ga4tp/?view_only=6e0f6b455db14bf79882c0f910277d96
25

26 253

27 28 254 **Results**

29 30 255 **Assessments of biological traits**

31 32 256 ***Sex/gender recognition***

33
34
35 257 The overall accuracy of voice-based sex/gender judgments was very high, averaging
36
37 258 99% for both types of speech stimuli (read and free speech) and for both sexes of vocalizer.
38
39
40 259 None of the predictors included in our models significantly predicted variance in sex rating
41
42 260 accuracy (see Table S1 in the supplementary materials). Accuracy in voice-based sex
43
44 261 recognition was, therefore, extremely high independent of method of data collection, speaker
45
46 262 age, rater age, and rater gender for both female and male vocalizers, or speech type. **Indeed,**
47
48 263 **the effect of speech type was very small, $b = -0.001$ CIs: $-0.01, 0.01, p = .781$ and $b < 0.001$**
49
50 264 **CIs: $-0.01, 0.01, p = .857$ for female and male vocalisers, respectively, indicating that listeners**
51
52
53 265 **could gauge sex extremally and equally well from both read and free speech.**

54 55 56 266 ***Age assessment***

57
58
59
60

1
2
3 267 For both male and female vocalizers, assessed age was predicted by the vocalizers
4
5 268 actual age (females: $b = 0.69, p < .001$, males: $b = 0.524, p < .001$), indicating an overall high
6
7 269 accuracy in listeners' voice-based ratings of age. For female vocalizers, assessed age was not
8
9 270 statistically significantly related to the speech type ($b = -0.50$ CIs: $-0.84, -0.06, p = .026$), with
10
11 271 read speech eliciting older judgements than free speech. The speech type did not however
12
13 272 predict *accuracy* in listeners' age assessments of neither female ($b = -0.03$, CIs: $-0.07, 0.002$,
14
15 273 $p = .062$) nor male vocalizers ($b = -0.03$, CIs: $-0.07, 0.003, p = .069$) as reflected by a lack of
16
17 274 interaction effects. In males, speech type also did not explain any variance in raters' age
18
19 275 assessments ($b = -0.01$ CIs: $-0.42, -0.44, p = .967$) (see Table S2 for outcomes). *Whether*
20
21 276 *participants completed the experiment in the lab or online did not significantly predict their*
22
23 277 *age assessments of either sex ($p = .318$ and $p = .720$ for female and male vocalizers,*
24
25 278 *respectively).*

279 ***Height assessment***

280 In both female and male vocalizers, actual height predicted perceived height but for
281 females this relationship was slightly below the threshold of statistical significance following
282 Bonferroni correction ($b = 0.09, p = .014$ in females and $b = 0.19, p < .001$ in males). For
283 female vocalizers, speech type did not predict listeners' assessments of speaker height ($b = -$
284 0.32 , CIs: $-0.71, 0.07, p = .109$, see Table S3 for detailed outcomes) and there was no
285 significant interaction between speech type and vocalizer's actual height on listeners'
286 assessments ($b = -0.09$, CIs: $-0.16, -0.013, p = .022$). In male vocalizers, no effect of speech
287 type was found ($b = 0.08$, CIs: $-0.36, 0.52, p = .728$), and there was no interaction between
288 actual and assessed height ($b = -0.02$, CIs: $-0.09, 0.06, p = .667$). *Whether participants*
289 *completed the experiment in the lab or online did not significantly predict their height*
290 *assessments of either sex ($p = .114$ and $p = .062$ for female and male vocalizers, respectively).*

291 ***Weight assessment***

1
2
3 292 For female vocalizers, the model revealed no significant main effect of female
4
5 293 vocalizers' actual weights on listeners' weight assessments following Bonferroni correction (b
6
7 294 = 0.07, $p = .018$). We found no effect of speech type on listener's assessments of weight ($b =$
8
9 295 0.02, CIs: -0.48, 0.45, $p = .948$), and no interaction between speech type and vocalizer actual
10
11 296 weight ($b = -0.03$, CIs: -0.07, 0.006, $p = .099$). In male vocalizers, the effect of speech type
12
13 297 was also not significant ($b = 0.05$, CIs: -0.47, 0.57, $p = .845$), nor was the effect of
14
15 298 participant's actual weight on their assessed weight ($p = .137$), with no interaction between
16
17 299 these two variables ($b = 0.02$, CIs: -0.02, 0.06 $p = .257$). Thus, these results suggest that the
18
19 300 type of speech stimulus did not influence listeners' weight judgements (see Table S4).
20
21 301 Whether participants completed the experiment in the lab or online did not significantly
22
23 302 predict their weight assessments of either sex ($p = .875$ and $p = .982$ for female and male
24
25 303 vocalizers, respectively).
26
27
28
29
30

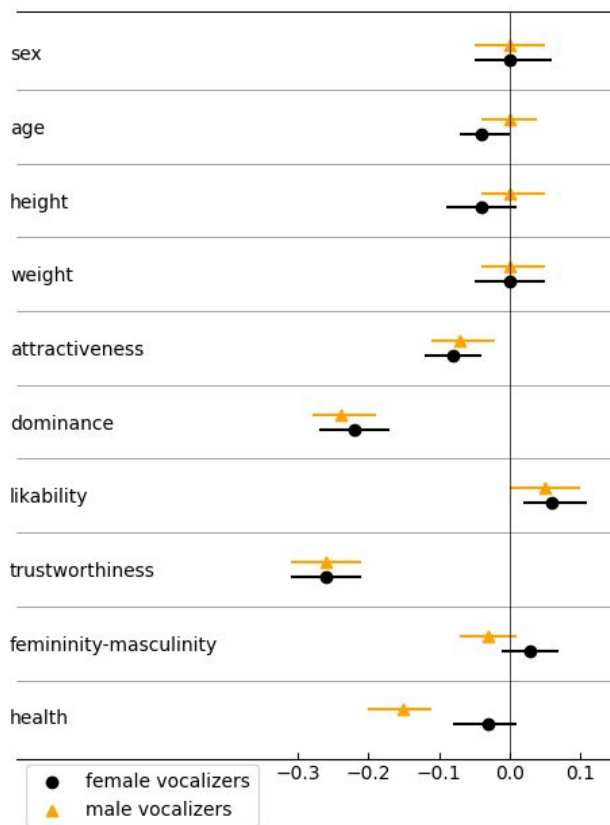
31 304 **Assessment of psychosocial traits**

32
33
34
35 305 Subsequent models examined listeners' assessments of socially relevant traits of
36
37 306 vocalizers including attractiveness, dominance, likeability, trustworthiness,
38
39 307 masculinity/femininity, and health. While many of these traits can be framed as both
40
41 308 psychological and biological to some degree, the key distinction for the purposes of this study
42
43 309 is that these traits were not defined by any objective ground-truth measure. Thus, while we
44
45 310 could test whether listeners' perceptions of these traits varied when judging free versus read
46
47 311 speech, we could not test whether speech type affected the accuracy of these assessments, as
48
49 312 we could for biological traits for which we had objective measures (sex, age, and body size).
50
51 313 The statistical models were thus composed of the same set of predictors (speech type,
52
53 314 vocalizer/rater age and sex), however we did not assess accuracy in listeners' assessments of
54
55 315 these psychosocial traits.
56
57
58
59
60

1
2
3 316 Speech type generally predicted differences in listeners' ratings of psychosocial traits.
4
5 317 Read speech elicited relatively higher ratings of attractiveness (vocalizer sex: females $b = -$
6
7 318 0.13, CIs: -0.20, -0.06, $p < .001$; males $b = -0.11$, CIs: -0.18, -0.04, $p = .003$; Table S5) and
8
9 319 dominance (females $b = -0.37$, CIs: -0.45, -0.29, $p < .001$; males $b = -0.39$, CIs: -0.47, -0.31, p
10
11 320 $< .001$; Table S6) and trustworthiness (females $b = -0.45$, CIs: -0.53, -0.36 $p < .001$; males $b =$
12
13 321 -0.44, CIs: -0.52, -0.36, $p < .001$ Table S8) than than did free speech. Likability ratings did
14
15 322 not vary significantly in females ($b = 0.099$, CIs: 0.024, 0.173, $p = .009$) and males ($b = 0.08$,
16
17 323 CIs: 0.01, 0.15, $p = .033$; Table S7). Masculinity-femininity ratings also did not vary between
18
19 324 speech types for either female ($b = 0.04$, CIs: -0.01, 0.10, $p = .146$) or male vocalizers ($b = -$
20
21 325 0.037, CIs: -0.09, 0.02, $p = .165$; Table S9), nor did health assessments of female vocalizers
22
23 326 ($b = -0.05$, CIs: -0.12, 0.02, $p = .152$), whereas the same male vocalizers producing read,
24
25 327 rather than free speech, were evaluated as healthier ($b = -0.25$, CIs: -0.32, -0.17, $p < .001$;
26
27 328 Table S10). In general, these differences were small.

28
29
30
31
32
33
34 329 In all cases, whether participants completed the experiment in the lab or online did not
35
36 330 affect the pattern of results, with the exception of ratings of men's health ($b = -0.20$, CIs: -
37
38 331 0.44, 0.03, $p < 0.001$). Therefore, we performed additional analyses for lab and online raters
39
40 332 with perceived health as an outcome variable (see Table S11). The effect of speech stimulus
41
42 333 type was statistically significant in lab ($b = -0.21$, CIs: -0.32, -0.09, $p < .001$) but not online
43
44 334 raters ($b = -0.03$, CIs: -0.14, 0.09, $p = .663$) with higher ratings observed for read speech (as
45
46 335 compared to free speech).

47
48
49
50
51
52
53
54
55
56
57
58
59
60



336

337 **Figure 1. Listeners' ratings of speaker traits based on read versus free speech.**

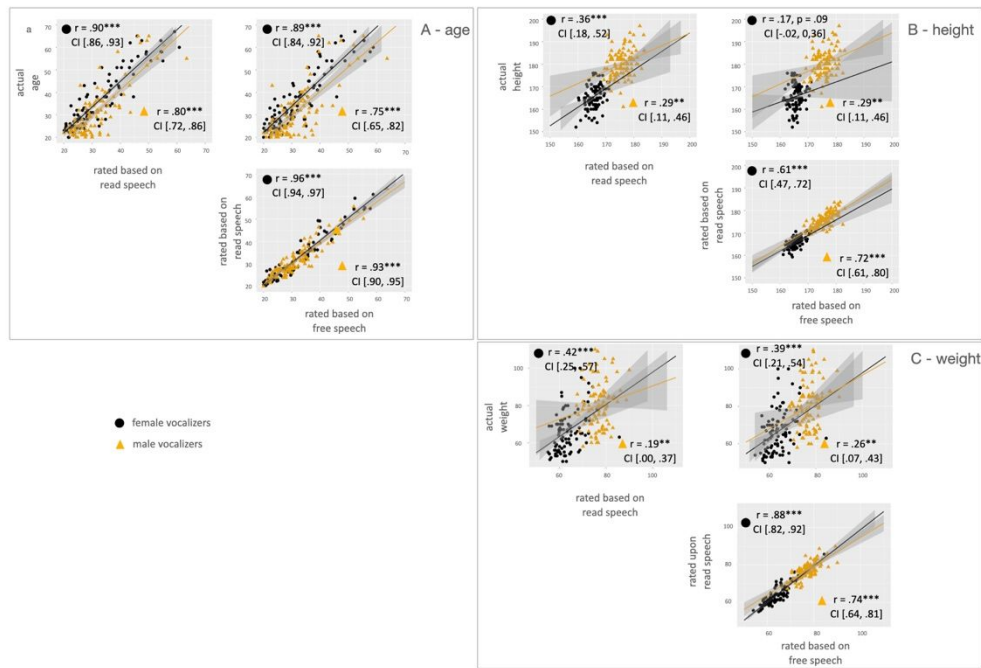
338 Standardized beta coefficients with 95% CIs representing the effects of free vs. read speech
 339 across all traits of interest (each coefficient is derived from a separate analysis) for female and
 340 male vocalizers. The plot illustrates that read speech elicited relatively higher ratings of
 341 attractiveness and dominance, whereas free speech elicited relatively higher ratings of
 342 likeability and trustworthiness.

343

344 **Additional analyses**

345 To corroborate our findings, in addition to the multilevel models reported above, we
 346 performed analyses at the vocalizer level ($n = 208$). Across all ratings for which we collected
 347 objective continuous measures of traits from vocalizers (i.e., actual age, height and weight),
 348 we conducted correlation analyses to illustrate accuracy in listeners' ratings by speech type

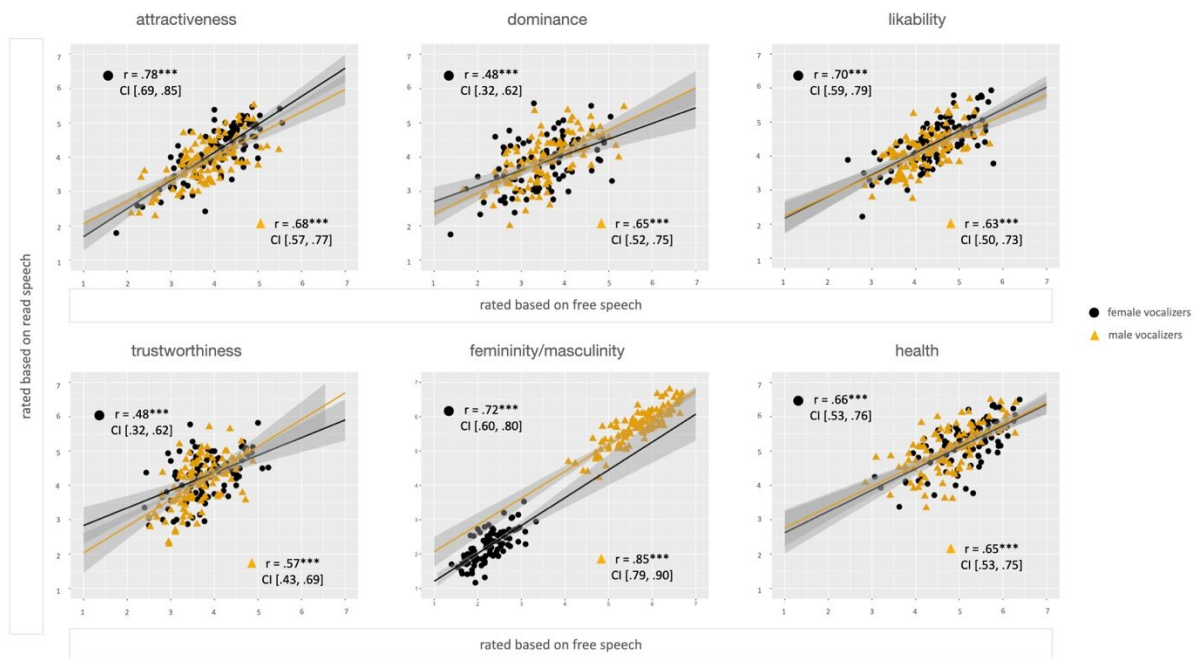
1
2
3 349 (i.e., we correlated assessed traits with objective measures of those traits for both read and
4
5 350 free speech). The results are presented in Figure 2, whereas in Figure 3 we further present
6
7 351 correlations between mean ratings of psychosocial traits based on read versus free speech. In
8
9 352 both sets of these regression analyses, the relationships are significant in all cases (at $p < .001$)
10
11 353 and range from *moderate* ($r = .48, p < .001$) in the case of dominance ratings of male
12
13 354 vocalizers, to *extremely strong* ($r = .96, p < .001$) in the case of age ratings for female
14
15 355 vocalizers (see Figures 2 and 3). These results demonstrate that listeners' voice-based ratings
16
17 356 of vocalizer traits correlate substantially between read and free speech produced by the same
18
19 357 vocalizers, indicating that vocalizers are perceived similarly regardless whether they are
20
21 358 reading a scripted passage or producing free speech about the weather.
22
23
24
25
26
27
28
29
30
31 360
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



361

362 **Figure 2. Relationships between actual objective measures and listeners' ratings of**
 363 **biological vocalizer traits for read versus free speech.** The upper rows in each panel (A, B,
 364 C) reflect correlations between objective measures of actual age/height/weight and listeners'
 365 mean ratings for these same traits, based on read versus free speech. The lower scatterplot in
 366 each panel represents the relationship between mean ratings from read and free speech for
 367 each biological trait. Each datapoint presents a single vocalizer ($n = 208$), wherein black
 368 circles represent female vocalizers and yellow diamonds represent male vocalizers. Grey
 369 shading represents 95% CIs. Mean ratings are derived from voice-based assessments given by
 370 4088 male and female listeners.

371



372

373 **Figure 3. Relationships in listeners' ratings of psychosocial traits based on read versus**

374 **free speech.** Relationships between listeners' ratings of psychosocial traits based on read

375 versus free speech were moderate to strong. Each datapoint represents a single vocalizer ($n =$

376 208), wherein black circles represent female vocalizers and yellow diamonds represent male

377 vocalizers. Grey shading represents 95% CIs. Mean ratings are derived from voice-based

378 assessments given by 4088 male and female listeners.

379

380

381 **Discussion**

382 In our species, as in many other terrestrial mammals, the voice is a crucial modality
383 for social communication. This study corroborates a large body of literature showing that
384 simply by hearing another person speak, even for a brief moment and largely regardless of
385 what they are saying, we can effectively gauge their sex, their general age and body size, and
386 we can make a rapid series of judgments regarding other psychological and social traits, such
387 as how attractive, dominant or trustworthy they are (Kreiman & Sidtis, 2011; Kamiloğlu &
388 Sauter, 2021; Pisanski and Bryant, 2019 for reviews). Here, we go one step further to also
389 show that these judgments are highly similar for the same speaker whether they are reading a
390 passage or producing free speech about the same neutral topic. In other words, a person who
391 is perceived as particularly dominant or as relatively young when reading is perceived very
392 similarly when speaking spontaneously, even when using different words, though this
393 depends to some extent on the trait being judged as described below. Listeners are also
394 similarly accurate in their ability to judge objective traits, such as sex, age and body size,
395 from read and free speech. Thus, despite some acoustic variation in how we humans talk
396 when we read, largely tied to breathing (Winkworth et al., 1994), this mild departure from our
397 natural spontaneous way of speaking does not profoundly affect listeners' judgments of us.
398 This is almost certainly due to intraindividual stability in underlying vocal parameters such as
399 fundamental and formant frequencies across different kinds of speech (see e.g., Pisanski et al.,
400 2016, 2021; Pisanski, Raine & Reby, 2020) that often explain the majority of variance in
401 listeners' voice-based judgments of speakers (Pisanski and Bryant, 2019 for review).

402 While the present study shows that voice-based assessments of biological traits,
403 including speaker sex, age, and body size, depend very little on whether a person is speaking
404 freely or reading a scripted text, one exceptions is age assessments of female vocalizers,
405 wherein listener's judged women as slightly older (< 1 year) when those women produced

1
2
3 406 read speech versus when they spoke freely about the weather, a relatively small difference in
4
5 407 ecological terms considering that our sample of vocalizers ranged from 18 to 67 years of age.
6
7 408 Moreover, accuracy in age assessments of female vocalizers was the same for read and free
8
9 409 speech.

10
11
12 410 Interestingly, we show that for assessments of some psychosocial traits (e.g.,
13
14 411 attractiveness, dominance, and trustworthiness), [listeners' judgments are slightly more](#)
15
16 412 [affected by whether speech is read or spontaneous compared to what we observed for](#)
17
18 413 [quantifiable biological traits](#). Indeed, we observed statistically significant differences between
19
20 414 listeners' ratings for read and free speech across most psychosocial traits and for both
21
22 415 vocalizer sexes, with the exceptions of masculinity-femininity assessments of males and
23
24 416 females and health assessments of females. [It should be noted, however, that the effect sizes](#)
25
26 417 [were small and not likely to be ecologically or socially meaningful](#).

27
28
29
30 418 Critically, using regression analyses, we show that listeners' ratings of both biological
31
32 419 and psychosocial traits share a high degree of variance between speech types for the same
33
34 420 speakers. Even in cases where the differences between read and free speech were significant,
35
36 421 ratings between speech types were moderately to strongly correlated for the same group of
37
38 422 vocalizers, [with at least 23% and up to 92% of the variance in ratings for one speech type](#)
39
40 423 [explained by ratings for the other](#). Nevertheless, the finding that assessments based on
41
42 424 different types of speech can differ is important to consider when comparing voice-based
43
44 425 ratings across studies that use different speech types, as well as in designing further research.
45
46 426 Regarding practical implications, such differences may be relevant in the social and media
47
48 427 spheres, for example in terms of how listeners may regard public figures or persons in
49
50 428 advertisements based on the perceived spontaneity of their speech.

51
52
53
54 429 Our study also corroborates previous studies showing that human listeners can judge
55
56 430 physical characteristics fairly reliably from the voice alone ([for reviews see Kamiloğlu &](#)
57
58
59
60

1
2
3 431 Sauter, 2021; Pisanski & Bryant, 2019). In our study with a large sample of over 4000
4
5 432 listeners, we have shown that sex/gender was assessed almost perfectly, age with high
6
7 433 accuracy, and body size still higher than chance. Weight ratings were the least accurate – in
8
9 434 particular, male weight was not accurately recognized by our listeners from the voice,
10
11 435 regardless of speech type.
12
13
14

15 436 **Limitations and future research recommendations**

16
17
18
19 437 Our research design does not allow us to draw direct conclusions regarding the
20
21 438 ‘accuracy’ of listeners’ judgments for psychosocial traits. We appreciate that many of these
22
23 439 traits, such as masculinity, femininity, attractiveness, dominance and health, have biological
24
25 440 components. Nevertheless, in this study, we did not have quantifiable, objective measures of
26
27 441 these traits, nor of likeability or trustworthiness. It is indeed difficult to quantify such traits
28
29 442 (see e.g., Bem, 1974; Rubenstein, 2005) compared to objective traits such as age or body size.
30
31 443 Our study also has technical limitations, for example, although online raters were instructed to
32
33 444 use high quality headphones, and headphone use was verified with hearing tests, we cannot be
34
35 445 certain that the quality of their headphones was comparable to that of those used by lab
36
37 446 participants. However, this method allowed us to increase our sample of respondents, and we
38
39 447 did not find systematic differences in our results between our lab and online samples. Finally,
40
41 448 as our sample was homogenous in terms of language, the study may be replicated in other
42
43 449 countries to assess its cross-cultural generalizability. In the future, it would also be
44
45 450 informative to examine the effect of speech type in more ecologically valid or multi-modal
46
47 451 conditions, i.e., where a speaker is seen (Groyecka et al., 2017), to test whether speech type
48
49 452 remains relevant when other social cues and modalities are available to the perceiver.
50
51
52
53
54
55
56

57 453 **Conclusion**

58
59
60

1
2
3 454 The human voice can be a critical social tool that informs listeners about biological and
4
5 455 psychosocial traits of human vocalizers. While past work has shown that some acoustic
6
7 456 properties of the voice, such as fundamental frequency (voice pitch), are relatively stable
8
9 457 across speech types (Pisanski et al., 2016, 2021), and that listeners' judgments also do not
10
11 458 vary substantially based on the duration or complexity of scripted speech (Makrholz et al.,
12
13 459 2018; Groyecka-Bernard et al. 2022), this study is among the first to compare listeners'
14
15 460 ratings of vocalizers based on read versus free speech, and for a broad range of biological and
16
17 461 psychosocial traits. Our results thus provide the most comprehensive evidence to date that
18
19 462 listeners' judgments of speakers based on vocal cues and the validity of those judgments also
20
21 463 vary to little yet observable extent with the type of speech stimulus (free speech vs read
22
23 464 speech). This indicates that while the same vocalizer might be perceived as, for example,
24
25 465 slightly more trustworthy when producing read than free speech, both those ratings will
26
27 466 correlate, often strongly, for any given speaker. Nevertheless, although differences in voice-
28
29 467 based assessments for free versus read speech are minor, they are worth acknowledging when
30
31 468 comparing across studies or designing future research paradigms.
32
33
34
35
36
37
38

469 **Declarations**

40
41
42 470 **Funding:** This work was supported by a grant from the National Science Centre (grant no.
43
44 471 2016/23/B/HS6/00771).

45
46
47 472 **Conflicts of interest/Competing interest:** The authors declare no conflicts of interest.

48
49 473 **Ethics approval:** Study protocols were accepted by the ethics committee at the Institute of
50
51 474 Psychology, University of [*masked for peer review*].

52
53
54 475 **Consent to participate:** Vocalizers and raters provided informed consent prior to study
55
56 476 inclusion.

57
58 477 **Consent for publication:** The manuscript has been read and approved by all authors.
59
60

1
2
3 478
4
5
6
7 479
8
9

10 480

References

11
12
13 481 Atkinson, J., Pipitone, R. N., Sorokowska, A., Sorokowski, P., Mberira, M., Bartels, A., &
14
15
16 482 Gallup Jr., G. G. (2012). Voice and Handgrip Strength Predict Reproductive Success
17
18 483 in a Group of Indigenous African Females. *PloS One*, 7(8): e41811.

19
20 484 Ari, N., & Ustazhanov, M. (2014, September). Matplotlib in python. In *2014 11th*

21
22 485 *International Conference on Electronics, Computer and Computation (ICECCO)*

23
24 486 (pp. 1-6). IEEE.

25
26
27 487 Aung, T. & Puts, D. (2020). Voice pitch: A window into the communication of social power,

28
29 488 *Current Opinion in Psychology*, 33, 154-161.

30
31 489 Ferdenzi, C., Patel, S., Mehu-Blantar, I., Khidasheli, M., Sander, D., & Delplanque, S. (2013).

32
33 490 Voice attractiveness: Influence of stimulus duration and type. *Behavior Research*

34
35 491 *Methods*, 45(2), 405–413.

36
37 492 Groyecka-Bernard, A., Pisanski, K., Frąckowiak, T., Kobylarek, A., Kupczyk, P.,

38
39 493 Oleszkiewicz, A., Sabiniewicz, A., Wróbel, M., & Sorokowski, P. (2022). Do voice-

40
41 494 based judgments of socially relevant speaker traits differ across speech

42
43 495 types?. *Journal of Speech, Language, and Hearing Research*, 65(10), 3674-3694.

44
45 496 Groyecka, A., Pisanski, K., Sorokowska, A., Havlíček, J., Karwowski, M., Puts, D., Roberts,

46
47 497 C. S., & Sorokowski, P. (2017). Attractiveness is multimodal: beauty is also in the

48
49 498 nose and ear of the beholder. *Frontiers in Psychology*, 8:778.

50
51 499 Jensen, A. R. (2001). Vocabulary and general intelligence. *Behavioral and Brain*

52
53 500 *Sciences*, 24(6), 1109-1110.

54
55
56
57
58
59
60

- 1
2
3 501 Kamiloğlu, R. G., & Sauter, D. A. (2021). Voice Production and Perception. In *Oxford*
4
5 502 *Research Encyclopedia of Psychology*.
6
7 503 Knowles, K. K., & Little, A. C. (2016). Vocal fundamental and formant frequencies affect
8
9
10 504 perceptions of speaker cooperativeness. *Quarterly Journal of Experimental*
11
12 505 *Psychology*, 69(9), 1657-1675.
13
14 506 Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative
15
16 507 importance of intra- and intersexual selection on human male sexually dimorphic
17
18 508 traits. *Evolution and Human Behavior*, 39(4), 424–436.
19
20 509 Kreiman, J. & Sidtis, D. 2011. *Foundation of voice studies: An interdisciplinary approach to*
21
22 510 *voice production and perception*. John Wiley & Sons.
23
24 511 Lavan, N. (2023). The time course of person perception From voices: A behavioral study.
25
26 512 *Psychological Science*, 09567976231161565.
27
28 513 Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, 21(4), R143-
29
30 514 R145.
31
32 515 Lüdecke, D., Ben-Shachar, M. S., Patil, I., & Makowski, D. (2020). Extracting, computing
33
34 516 and exploring the parameters of statistical models using R. *Journal of Open Source*
35
36 517 *Software*, 5(53), 2445.
37
38 518 Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are
39
40 519 correlated across differing content and stimulus type. *PloS One*, 13(10), e0204991.
41
42 520 Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch and the labor market
43
44 521 success of male chief executive officers. *Evolution and Human Behavior*, 34(4), 243-
45
46 522 248.
47
48 523 McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality
49
50 524 impressions from brief novel voices. *PloS one*, 9(3), e90779.
51
52
53
54
55
56
57
58
59
60

- 1
2
3 525 Painter, D., Fasoli, F., & Sulpizio, S. (2021). The impact of stimuli length and analytic
4
5 526 method on auditory 'Gaydar' research. *Journal of Voice*.
- 6
7 527 Pipitone, N.R., & Gallup Jr, G. G. (2008). Women's voice attractiveness varies across the
8
9 528 menstrual cycle. *Evolution and Human Behavior*, 29(4), 268-274.
- 10
11 529 Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. *The oxford handbook*
12
13 530 *of voice studies*, 269-300.
- 14
15 531 Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P. W., Fink,
16
17 532 B., DeBruine, L. M., Jones, B. C., & Feinberg, D. R. (2014). Vocal indicators of
18
19 533 body size in men and women: a meta-analysis. *Animal Behaviour*, 95, 89-99.
- 20
21 534 Pisanski, K., Nowak, J., & Sorokowski, P. (2016). Individual differences in cortisol stress
22
23 535 response predict increases in voice pitch during exam stress. *Physiology &*
24
25 536 *Behavior*, 163, 234-238.
- 26
27 537 Pisanski, K., Groyecka-Bernard, A., & Sorokowski, P. (2021). Human voice pitch measures
28
29 538 are robust across a variety of speech recordings: methodological and theoretical
30
31 539 implications. *Biology Letters*, 17(9), 20210356.
- 32
33 540 Pisanski, K., Raine, J., & Reby, D. (2020). Individual differences in human voice pitch are
34
35 541 preserved from speech to screams, roars and pain cries. *Royal Society open*
36
37 542 *science*, 7(2), 191642.
- 38
39 543 Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as
40
41 544 dominance signals: vocal fundamental and formant frequencies influence dominance
42
43 545 attributions among men. *Evolution and Human Behavior*, 28(5), 340-344.
- 44
45 546 Rubenstein, A. J. (2005). Variation in perceived attractiveness: Differences between dynamic
46
47 547 and static faces. *Psychological Science*, 16(10), 759-762.
- 48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 548 Schild, C., Aung, T., Kordsmeyer, T. L., Cardenas, R. A., Puts, D. A., & Penke, L. (2020).
4
5 549 Linking human male vocal parameters to perceptions, body morphology, strength
6
7 550 and hormonal profiles in contexts of sexual selection. *Scientific Reports*, *10*(1), 1-16.
8
9
10 551 Schild, C., Stern, J., Zettler, I. (2020). Linking men's voice pitch to actual and perceived
11
12 552 trustworthiness across domains. *Behavioral Ecology*, *31*(1), 164–175.
13
14 553 Sorokowski, P., Groyecka-Bernard, A., & Luty, J. (2022). Methodological aspects of using
15
16 554 different lengths stimuli in voice research: short review. *Proceedings of Meetings on*
17
18 555 *Acoustics*, *46*, 050001.
19
20
21 556 Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A.,
22
23 557 Kowal, M., Borkowska, B., & Pisanski, K. (2019). Voice of authority: professionals
24
25 558 lower their vocal frequencies when giving expert advice. *Journal of Nonverbal*
26
27 559 *Behavior*, *43*(2), 257-269.
28
29
30 560 Székely, É., Henter, G. E., Beskow, J., & Gustafson, J. (2020). Breathing and speech planning
31
32 561 in spontaneous speech synthesis. *ICASSP 2020-2020 IEEE International Conference*
33
34 562 *on Acoustics, Speech and Signal Processing (ICASSP)*, 7649-7653.
35
36
37 563 Székely, É., Henter, G. E., Beskow, J., & Gustafson, J. (2019). Spontaneous Conversational
38
39 564 Speech Synthesis from Found Data. *Interspeech*, 4435-4439.
40
41
42 565 Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch
43
44 566 influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210-216.
45
46
47 567 Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices.
48
49 568 *Journal of the Acoustical Society of America*, *85*(4), 1699–1707.
50
51 569 Wester, M., Watts, O., & Henter, G. E. (2016). Evaluating comprehension of natural and
52
53 570 synthetic conversational speech. *Speech Prosody*, *8*, 736–740.
54
55
56 571 Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. *Create elegant data*
57
58 572 *visualisations using the grammar of graphics. Version*, *2*(1), 1-189.
59
60

1
2
3 573 Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency
4
5 574 in speech breathing during reading: Lung volumes, speech intensity, and linguistic
6
7 575 factors. *Journal of Speech, Language, and Hearing Research*, 37(3), 535-556.
8
9

10 576

11
12 577
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Review Only

Supplementary Materials

Table S1

Model with sex accuracy rating an outcome

Female vocalizers						
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals		
				Lower	Upper	
Fixed effects						
Intercept	0.987	0.003	<.001	0.980	0.993	
Stimulus type (0-read, 1-free)	0.001	0.003	0.781	-0.006	0.007	
Vocalizer's properties						
Age	0.000	0.000	0.492	0.000	0.000	
Rater's properties						
Sex (0-F, 1-M)	-0.008	0.006	0.236	-0.020	0.004	
Age	0.000	0.000	0.221	0.000	0.000	
Method (0 – online, 1 – lab)	-0.012	0.007	0.098	-0.026	0.002	
Random effects						
	Variance		ICC			
Rater's ID	0.002		0.1700			
Vocalizer's ID	0.000		0.0172			
Residual	0.012					
N _{raters} =429, N _{vocalizers} =101, N _{observations} =4290						
Model parameter						
BIC	-6125.086					
Male vocalizers						
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals		
				Lower	Upper	
Fixed effects						

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

1						
2						
3	Intercept	0.007	0.002	<.001	0.003	0.012
4	Stimulus type (0-read, 1-free)	0.000	0.003	0.857	-0.005	0.006
5	Vocalizer's properties					
6	Age	0.000	0.000	0.917	-0.000	0.000
7	Rater's properties					
8	Sex (0-F, 1-M)	0.007	0.004	0.116	-0.002	0.015
9	Age	0.000	0.000	0.202	-0.000	0.001
10						
11						
12	Method (0 – online, 1 – lab)	0.008	0.005	0.128	-0.002	0.017
13						
14	Random effects					
15		Variance		ICC		
16	Rater's ID	0.001		0.121		
17	Vocalizer's ID	0.000		0.002		
18	Residual	0.007				
19	N _{raters} =429, N _{vocalizers} =106, N _{observations} =4210					
20						
21	Model parameter					
22	BIC	-8287.473				
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2

Model with rated age as outcome

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	31.435	0.465	< .001	30.522	32.347
Stimulus type (0-read, 1-free)	-0.499	0.224	0.026	-0.938	-0.060
Vocalizer's properties					
Age	0.693	0.033	< .001	0.629	0.757
Vocalizer's age*Stimulus Type	-0.033	0.018	0.062	-0.068	0.002
Rater's properties					
Sex (0-F, 1-M)	-1.123	0.489	0.022	-2.081	-0.166
Age	0.021	0.023	0.351	-0.023	0.066
Method (0 – online, 1 – lab)	0.584	0.584	0.318	-0.561	1.729
Random effects					
	Variance		ICC		
Rater's ID	17.817		0.260		
Vocalizer's ID	15.666		0.236		
Residual	50.809				
N _{raters} =405, N _{vocalizers} =101, N _{observations} =4047					
Model parameter					
BIC	28327.323				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	31.346	0.519	< .001	30.329	32.363

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

1						
2						
3	Stimulus type (0-read, 1-free)	0.009	0.219	0.967	-0.420	0.438
4	Vocalizer's properties					
5	Age	0.524	0.041	<.001	0.444	0.603
6	Vocalizer's age*Stimulus Type	-0.034	0.019	0.069	-0.071	0.003
7	Rater's properties					
8	Sex (0-F, 1-M)	-0.774	0.506	0.127	-1.766	0.218
9	Age	0.067	0.024	0.005	0.021	0.113
10						
11						
12	Method (0 – online, 1 – lab)	-0.217	0.606	0.720	-1.405	0.971
13						
14	Random effects	Variance	SD	ICC		
15	Rater's ID	19.715		0.292		
16	Vocalizer's ID	21.546		0.310		
17	Residual	47.857				
18						
19	$N_{\text{raters}}=405, N_{\text{vocalizers}}=101, N_{\text{observations}}=4003$					
20	Model parameter					
21	BIC	27879.172				
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S3*Model with rated height as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	165.580	0.320	< .001	164.954	166.207
Stimulus type (0-read, 1-free)	-0.320	0.199	0.109	-0.711	0.071
Vocalizer's properties					
Age	-0.033	0.016	0.037	-0.064	-0.002
Height	0.094	0.038	0.014	0.021	0.168
Vocalizer's height*Stimulus type	-0.086	0.038	0.022	-0.160	-0.013
Rater's properties					
Sex (0-F, 1-M)	0.592	0.546	0.279	-0.479	1.662
Age	-0.021	0.023	0.355	-0.067	0.024
Method (0 – online, 1 – lab)	1.051	0.664	0.114	-0.250	2.353
Random effects					
	Variance		ICC		
Rater's ID	22.310		0.383		
Vocalizer's ID	2.393		0.062		
Residual	35.976				
N _{raters} =362, N _{vocalizers} =101, N _{observations} =3619					
Model parameter					
BIC	24163.778				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	175.616	0.432	< .001	174.769	176.463

1						
2						
3	Stimulus type (0-read, 1-free)	0.078	0.224	0.728	-0.361	0.517
4	Vocalizer's properties					
5	Age	-0.022	0.022	0.314	-0.065	0.021
6	Height	0.191	0.041	<.001	0.111	0.272
7	Vocalizer's height*Stimulus type	-0.016	0.038	0.667	-0.090	0.057
8	Rater's properties					
9	Sex (0-F, 1-M)	0.441	0.735	0.548	-0.999	1.881
10	Age	-0.073	0.031	0.020	-0.134	-0.012
11						
12						
13	Method (0 – online, 1 – lab)	1.674	0.894	0.062	-0.078	0.062
14	Random effects	Variance		ICC		
15	Rater's ID	42.454		0.487		
16	Vocalizer's ID	4.752		0.096		
17	Residual	44.703				
18						
19	$N_{\text{raters}}=362, N_{\text{vocalizers}}=106, N_{\text{observations}}=3666$					
20	Model parameter					
21	BIC	24759.689				
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S4*Model with rated weight as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	63.456	0.423	< .001	62.627	64.285
Stimulus type (0-read, 1-free)	-0.015	0.237	0.948	-0.480	0.449
Vocalizer's properties					
Age	0.242	0.028	< .001	0.187	0.297
Weight	0.073	0.030	0.018	0.014	0.132
Vocalizer's weight*Stimulus type	-0.033	0.020	0.099	-0.072	0.006
Rater's properties					
Sex (0-F, 1-M)	0.357	0.582	0.541	-0.785	1.498
Age	-0.080	0.028	0.005	-0.135	-0.024
Method (0 – online, 1 – lab)	-0.118	0.749	0.875	-1.587	1.351
Random effects					
	Variance		ICC		
Rater's ID	24.449		0.452		
Vocalizer's ID	8.432		0.113		
Residual	51.284				
N _{raters} =365, N _{vocalizers} =101, N _{observations} =3650					
Model parameter					
BIC	25650.340				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	77.210	0.511	< .001	76.210	78.211
Stimulus type (0-read, 1-free)	0.052	0.264	0.845	-0.466	0.569

1						
2						
3	Vocalizer's properties					
4	Age	0.163	0.028	<.001	0.107	0.218
5	Weight	0.038	0.025	0.137	-0.012	0.087
6	Vocalizer's weight*Stimulus type	0.023	0.020	0.257	-0.017	0.062
7						
8	Rater's properties					
9	Sex (0-F, 1-M)	-1.706	0.813	0.036	-3.298	-0.113
10	Age	-0.059	0.039	0.131	-0.136	0.018
11						
12	Method (0 – online, 1 – lab)	0.023	1.044	0.982	-2.024	2.070
13						
14	Random effects		Variance		ICC	
15	Rater's ID		51.405		0.453	
16	Vocalizer's ID		7.912		0.113	
17	Residual		62.117			
18						
19	$N_{\text{raters}}=365, N_{\text{vocalizers}}=105, N_{\text{observations}}=3568$					
20	Model parameter					
21	BIC		25919.972			
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S5*Model with rated attractiveness as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.133	0.076	< .001	3.985	4.282
Stimulus type (0-read, 1-free)	-0.128	0.036	< .001	-0.199	-0.057
Vocalizer's properties					
Age	-0.029	0.005	< .001	-0.039	-0.019
Rater's properties					
Sex (0-F, 1-M)	-0.001	0.094	0.042	-0.184	0.017
Age	0.008	0.004		0.000	
			0.172		0.064
Method (0 – online, 1 – lab)	-0.148	0.108	< .001	-0.360	4.282
Random effects					
	Variance		ICC		
Rater's ID	0.758		0.346		
Vocalizer's ID	0.360		0.200		
Residual	1.436				
N _{raters} =435, N _{vocalizers} =101, N _{observations} =4350					
Model parameter					
BIC	15048.101				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.844	0.075	< .001	3.697	3.990
Stimulus type (0-read, 1-free)	-0.108	0.036	0.003	-0.178	-0.037

1						
2						
3	Vocalizer's properties					
4	Age	-0.011	0.005	0.033	-0.021	-0.001
5	Rater's properties					
6	Sex (0-F, 1-M)	-0.465	0.099	<.001	-0.660	-0.270
7	Age	0.003	0.004	0.490	-0.006	0.012
8						
9						
10	Method (0 – online, 1 – lab)	-0.115	0.115	0.318	-0.340	0.110
11						
12	Random effects		Variance		ICC	
13	Rater's ID		0.877		0.388	
14	Vocalizer's ID		0.327		0.191	
15	Residual		1.385			
16						
17	$N_{\text{raters}}=435, N_{\text{vocalizers}}=106, N_{\text{observations}}=4270$					
18						
19	Model parameter					
20	BIC		14696.740			
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S6*Model with rated dominance as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.640	0.077	< .001	3.489	3.791
Stimulus type (0-read, 1-free)	-0.373	0.041	< .001	-0.454	-0.293
Vocalizer's properties					
Age	0.001	0.005	0.845	-0.009	0.011
Rater's properties					
Sex (0-F, 1-M)	0.013	0.096	0.894	-0.175	0.200
Age	0.011	0.004	0.013	0.002	0.019
Method (0 – online, 1 – lab)	0.099	0.113	0.381	-0.122	0.320
Random effects					
	Variance		ICC		
Rater's ID	0.714		0.288		
Vocalizer's ID	0.369		0.172		
Residual	1.768				
N _{raters} =415, N _{vocalizers} =101, N _{observations} =4150					
Model parameter					
BIC	15123.168				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.717	0.082	< .001	3.557	3.877
Stimulus type (0-read, 1-free)	-0.393	0.040	< .001	-0.471	-0.314

1						
2						
3	Vocalizer's properties					
4	Age	0.006	0.006	0.346	-0.006	0.018
5	Rater's properties					
6	Sex (0-F, 1-M)	0.102	0.096	0.289	-0.086	0.290
7	Age	0.004	0.004	0.399	-0.005	0.012
8						
9						
10	Method (0 – online, 1 – lab)	0.053	0.113	0.643	-0.170	0.275
11						
12	Random effects		Variance		ICC	
13	Rater's ID		0.730		0.305	
14	Vocalizer's ID		0.457		0.216	
15	Residual		1.660			
16						
17	$N_{\text{raters}}=415, N_{\text{vocalizers}}=106, N_{\text{observations}}=4096$					
18						
19	Model parameter					
20	BIC		14646.311			
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S7*Model with rated likability as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.434	0.073	< .001	4.292	4.577
Stimulus type (0-read, 1-free)	0.099	0.038	0.009	0.024	0.173
Vocalizer's properties					
Age	-0.017	0.005	< .001	-0.026	-0.008
Rater's properties					
Sex (0-F, 1-M)	-0.117	0.092	0.203	-0.297	0.063
Age	0.006	0.004	0.174	-0.003	0.015
Method (0 – online, 1 – lab)	0.141	0.121	0.243	-0.096	0.378
Random effects					
	Variance		ICC		
Rater's ID	0.711		0.316		
Vocalizer's ID	0.312		0.168		
Residual	1.542				
N _{raters} =430, N _{vocalizers} =101, N _{observations} =4300					
Model parameter					
BIC	15118.559				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.249	0.065	< .001	4.121	4.377
Stimulus type (0-read, 1-free)	0.078	0.037	0.033	0.006	0.150

1						
2						
3	Vocalizer's properties					
4	Age	-0.001	0.004	0.893	-0.009	0.008
5	Rater's properties					
6	Sex (0-F, 1-M)	-0.223	0.090	0.014	-0.400	-0.047
7	Age	-0.001	0.004	0.843	-0.009	0.008
8						
9						
10	Method (0 – online, 1 – lab)	-0.026	0.118	0.826	-0.258	0.206
11						
12	Random effects		Variance		ICC	
13	Rater's ID		0.688		0.327	
14	Vocalizer's ID		0.227		0.138	
15	Residual		1.418			
16						
17	$N_{\text{raters}}=430, N_{\text{vocalizers}}=106, N_{\text{observations}}=4234$					
18						
19	Model parameter					
20	BIC		14835.407			
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S8

Model with rated trustworthiness as outcome

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.029	0.071	<.001	3.889	4.169
Stimulus type (0-read, 1-free)	-0.447	0.044	<.001	-0.533	-0.361
Vocalizer's properties					
Age	-0.002	0.004	0.677	-0.010	0.007
Rater's properties					
Sex (0-F, 1-M)	-0.083	0.103	0.422	-0.285	0.119
Age	0.005	0.005	0.330	-0.005	0.014
Method (0 – online, 1 – lab)	0.029	0.126	0.817	-0.217	0.276
Random effects					
	Variance		ICC		
Rater's ID	0.835		0.304		
Vocalizer's ID	0.234		0.109		
Residual	1.909				
N _{raters} =399, N _{vocalizers} =101, N _{observations} =3990					
Model parameter					
BIC	14835.407				

Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.852	0.068	<.001	3.718	3.986
Stimulus type (0-read, 1-free)	-0.440	0.043	<.001	-0.524	-0.356
Vocalizer's properties					

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

1						
2						
3	Age	0.005	0.004	0.291	-0.004	0.013
4	Rater's properties					
5	Sex (0-F, 1-M)	-0.046	0.102	0.648	-0.246	0.153
6	Age	0.002	0.005	0.661	-0.007	0.011
7						
8						
9	Method (0 – online, 1 – lab)	0.152	0.124	0.219	-0.090	0.395
10						
11	Random effects		Variance		ICC	
12	Rater's ID		0.814		0.313	
13	Vocalizer's ID		0.205		0.103	
14	Residual		1.787			
15						
16	$N_{\text{raters}}=399, N_{\text{vocalizers}}=106, N_{\text{observations}}=3888$					
17						
18	Model parameter					
19	BIC		14223.663			
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S9*Model with rated femininity-masculinity as outcome. Higher score indicated higher masculinity*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	2.228	0.061	<.001	2.109	2.347
Stimulus type (0-read, 1-free)	0.040	0.028	0.146	-0.014	0.095
Vocalizer's properties					
Age	0.008	0.003	0.003	0.003	0.014
Rater's properties					
Sex (0-F, 1-M)	-0.169	0.105	0.110	-0.375	0.038
Age	-0.007	0.005	0.162	-0.016	0.003
Method (0 – online, 1 – lab)	0.140	0.128	0.275	-0.111	0.390
Random effects					
	Variance		ICC		
Rater's ID	1.004		0.562		
Vocalizer's ID	0.098		0.112		
Residual	0.783				
N _{raters} =407, N _{vocalizers} =101, N _{observations} =4070					
Model parameter	11886.445				
BIC					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.686	0.064	<.001	5.561	5.811
Stimulus type (0-read, 1-free)	-0.037	0.027	0.165	-0.089	0.015
Vocalizer's properties					

Age	0.009	0.004	0.058	-0.000	0.017
Rater's properties					
Sex (0-F, 1-M)	-0.051	0.083	0.536	-0.214	0.111
Age	-0.000	0.004	0.951	-0.008	0.007
Method (0 – online, 1 – lab)					
	-0.120	0.101	0.233	-0.317	0.077
Random effects					
	Variance		ICC		
Rater's ID	0.596		0.456		
Vocalizer's ID	0.250		0.260		
Residual	0.711				
N _{raters} =407, N _{vocalizers} =106, N _{observations} =3995					
Model parameter					
BIC	11238.870				

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

Table S10*Model with rated health as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.187	0.072	<.001	5.046	5.328
Stimulus type (0-read, 1-free)	-0.052	0.036	0.152	-0.123	0.019
Vocalizer's properties					
Age	-0.020	0.005	<.001	-0.029	-0.011
Rater's properties					
Sex (0-F, 1-M)	-0.035	0.091	0.702	-0.214	0.144
Age	0.006	0.004	0.193	-0.003	0.014
Method (0 – online, 1 – lab)	-0.041	0.113	0.716	-0.262	0.180
Random effects					
	Variance		ICC		
Rater's ID	0.730		0.334		
Vocalizer's ID	0.307		0.174		
Residual	1.459				
N _{raters} =441, N _{vocalizers} =101, N _{observations} =4410					
Model parameter					
BIC	15196.876				
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.984	0.076	<.001	4.835	5.132
Stimulus type (0-read, 1-free)	-0.248	0.037	<.001	-0.321	-0.174
Vocalizer's properties					

1						
2						
3	Age	-0.007	0.005	0.177	-0.018	0.003
4	Rater's properties					
5	Sex (0-F, 1-M)	0.028	0.098	0.777	-0.164	0.219
6	Age	0.001	0.005	0.913	-0.008	0.009
7						
8						
9	Method (0 – online, 1 – lab)	-0.203	0.120	<.001	-0.438	0.032
10						
11	Random effects		Variance		ICC	
12	Rater's ID		0.842		0.358	
13	Vocalizer's ID		0.351		0.189	
14	Residual		1.510			
15						
16	$N_{\text{raters}}=441, N_{\text{vocalizers}}=106, N_{\text{observations}}=4306$					
17						
18	Model parameter					
19	BIC		15143.808			
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S11*Model with rated health as outcome for male vocalizers*

Online raters					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.087	0.094	<.001	4.903	5.271
Stimulus type (0-read, 1-free)	-0.207	0.059	<.001	-0.322	-0.092
Vocalizer's properties					
Age	-0.008	0.005	0.147	-0.018	0.003
Rater's properties					
Sex (0-F, 1-M)	-0.187	0.160	0.244	-0.501	0.127
Age	0.003	0.007	0.679	-0.011	0.017
Random effects					
	Variance		ICC		
Rater's ID	0.890		0.372		
Vocalizer's ID	0.252		0.143		
Residual	1.501				
N _{raters} =178, N _{vocalizers} =106, N _{observations} =1740					
Model parameter					
BIC	6187.296				
Lab raters					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.294	0.092	<.001	5.114	5.474
Stimulus type (0-read, 1-free)	-0.025	0.057	0.663	-0.136	0.087
Vocalizer's properties					
Age	-0.013	0.005	0.009	-0.023	-0.003
Rater's properties					

1						
2						
3	Sex (0-F, 1-M)	-0.196	0.152	0.197	-0.494	0.101
4	Age	0.007	0.007	0.305	-0.006	0.021
5						
6	Random effects		Variance		ICC	
7	Rater's ID		0.793		0.356	
8	Vocalizer's ID		0.273		0.160	
9	Residual		1.435			
10						
11	N _{raters} =178, N _{vocalizers} =101, N _{observations} =1786					
12						
13	Model parameter					
14	BIC		6239.034			
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

Table S12

Actual self-reported properties of the Vocalizers

	Age (years)			Height (cm)			Weight (kg)		
	M (SE)	Min	Max	M (SE)	Min	Max	M(SE)	Min	Max
Females	35.39 (12.68)	19	67	165.65 (5.36)	152	176	65.62 (11.62)	45	100
Males	30.40 (11.51)	18	65	180.20 (6.14)	161	197	81.03 (13.24)	58	130
Total	32.83 (12.32)	18	67	173.16 (9.29)	152	197	72.51 (14.64)	45	130

For Review Only

Table S13.*Properties of the vocalizers rated by raters depending on stimulus type.*

Estimated	Sex accuracy	Age	Height (kg)	Weight (kg)	Attractiveness	Dominance	Likability	Trustworthiness	Femininity- Masculinity	Health
Female vocalizers										
Read speech	99% (0.004)	31.68 (0.48)	165.74 (0.34)	63.4 (0.42)	4.20 (0.08)	3.83 (0.08)	4.39 (0.08)	4.26 (0.08)	2.21 (0.06)	5.21 (0.07)
Free speech	99% (0.004)	31.19 (0.48)	165.42 (0.34)	63.4 (0.42)	4.07 (0.08)	3.45 (0.08)	4.48 (0.08)	3.81 (0.08)	2.25 (0.06)	5.16 (0.07)
Male vocalizers										
Read speech	99% (0.002)	31.34 (0.53)	175.58 (0.44)	77.18 (0.50)	3.90 (0.10)	3.92 (0.08)	4.21 (0.07)	4.07 (0.07)	5.71 (0.07)	5.11 (0.08)
Free speech	99% (0.002)	31.35 (0.53)	175.66 (0.44)	77.24 (0.50)	3.79 (0.09)	3.53 (0.08)	4.29 (0.07)	3.63 (0.07)	5.67 (0.07)	4.84 (0.08)

*Note. Estimated marginal means are calculated with control for all covariates from the models. Standard errors are presented in parentheses. The values that differed significantly are marked as **bolded**.*

1
2
3 Repeated analyses were conducted, controlling for stimulus length. The obtained results are described below.
4 All tables for these analyses are labeled as S2.1, S2.2, S2.3, etc., and are located following the textual
5 description of the results.
6
7

8 9 **Statistical analyses**

10
11 To examine the effect of speech stimulus type on perceived traits of the vocalizers, as well as on the accuracy of listeners' ratings of
12 biological traits, we proceeded with a series of Linear Mixed Models (using Restricted Maximum Likelihood estimator) with responses clustered
13 both within speakers and raters. Because of sizeable sexual dimorphism in the acoustic properties of the human voice (Titze, 1989), we
14 conducted separate models for male and female vocalizers. Most models included ratings of a given trait as the outcome variable, speech type
15 (categorical variable coded 0 – free, 1 – read, please note that this is the opposite encoding from the main analysis described in the main article),
16 method of data collection (categorical variable coded 0 – online, 1 – lab), stimulus duration, vocalizer age, rater gender (categorical variable
17 coded 0-F, 1-M) and rater age as fixed variables, and vocalizers' and raters' coded identities as random effects. In the model examining the effect
18 of speech type on the accuracy of sex/gender ratings, we entered accuracy in sex assessment as the outcome variable (0-incorrect, 1-correct).
19 Additionally, in the models with age, height, and weight as outcomes, we included the actual height or weight of the speakers (as the relationship
20 between actual and rated age reflects the accuracy of the ratings) and age/height/weight*speech type as interaction effects, to test whether the
21 accuracy of listeners' ratings differs between read and free speech. All continuous predictors were grand-mean centred. All model results are
22 presented in Tables S2.1-S2.10 in the supplementary materials. For means and standard deviations of actual properties and rated properties (by
23 speech type), see Tables S2.11 and S2.12, respectively, in the supplementary materials. Voice samples for the current study were recorded from
24 198 adult men and women ($M_{\text{age}} = 33.03$, $SD_{\text{age}} = 12.46$; 46% women), For technical reasons, 10 recordings whose votes could not be estimated
25 were not included in the analysis.
26
27

28
29 Due to multiple comparisons, we performed Bonferroni correction and set the alpha level to 0.005 (0.05/10). Estimates presented in the
30 main text were not standardized. For standardized betas see Figure 1. All analyses and figures were performed using Jamovi, R using packages
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

ggplot2 (Wickham et al., 2016) and parameters (Lüdecke, 2020) and Python using matplotlib (Ari, 2014) package. Raw data and codebooks for Jamovi and scripts can be retrieved from: https://osf.io/ga4tp/?view_only=6e0f6b455db14bf79882c0f910277d96

Results

Assessments of biological traits

Sex/gender recognition

The overall accuracy of voice-based sex/gender judgments was very high, averaging 99% for both types of speech stimuli (read and free speech) and for both sexes of vocalizer. None of the predictors included in our models significantly predicted variance in sex rating accuracy (see Table S2.1 in the supplementary materials). Accuracy in voice-based sex recognition was, therefore, extremely high independent of method of data collection, speaker age, rater age, and rater gender for both female and male vocalizers, or speech type. Indeed, the effect of speech type was very small, $b = 0.001$ CIs: -0.02. 0.021, $p = .893$ and $b = 0.002$ CIs: -0.01. 0.01, $p = .766$ for female and male vocalisers, respectively, indicating that listeners could gauge sex extremely and equally well from both read and free speech.

Age assessment

For both male and female vocalizers, assessed age was predicted by the vocalizers actual age (females: $b = 0.69$, $p < .001$, males: $b = 0.55$, $p < .001$), indicating an overall high accuracy in listeners' voice-based ratings of age. For female vocalizers, assessed age was not statistically significantly related to the speech type ($b = -0.31$ CIs: -1.83. 1.21, $p = .69$). The speech type did not however predict *accuracy* in listeners' age assessments of neither female ($b = -0.03$, CIs: -0.07, 0.002, $p = .062$) nor male vocalizers ($b = -0.03$, CIs: -0.001, 0.07, $p = .055$) as

1
2
3 reflected by a lack of interaction effects. In males, speech type also did not explain any variance in raters' age assessments ($b = -0.49$ CIs: $-1.90, -$
4 $0.93, p = .499$) (see Table S2.2 for outcomes). Whether participants completed the experiment in the lab or online did not significantly predict
5 their age assessments of either sex ($p = .333$ and $p = .489$ for female and male vocalizers, respectively). Stimulus duration also did not predict
6 age assessments ($p = .300$ and $p = .494$ for female and male vocalizers, respectively).
7
8
9
10
11

12 ***Height assessment***

14
15 In both female and male vocalizers, actual height predicted perceived height but for females this relationship was slightly below the
16 threshold of statistical significance following Bonferroni correction ($b = 0.10, p = .012$ in females and $b = 0.21, p < .001$ in males). For female
17 vocalizers, speech type did not predict listeners' assessments of speaker height ($b = 1.57, \text{CIs: } 0.39, 2.75, p = .009$, see Table S2.3 for detailed
18 outcomes) and there was no significant interaction between speech type and vocalizer's actual height on listeners' assessments ($b = -0.09, \text{CIs: } -$
19 $0.01, 0.016, p = .02$). In male vocalizers, no effect of speech type was found ($b = -0.50, \text{CIs: } -1.79, 0.80, p = .45$), and there was no interaction
20 between actual and assessed height ($b = 0.03, \text{CIs: } -0.06, 0.12, p = .57$). Whether participants completed the experiment in the lab or online did
21 not significantly predict their height assessments of either sex ($p = .12$ and $p = .129$ for female and male vocalizers, respectively). Stimulus
22 duration also did not significantly predict height assessments but in females this relationship was slightly below the threshold of statistical
23 significance ($p = .024$ and $p = .522$ for female and male vocalizers, respectively).
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

Weight assessment

For female vocalizers, the model revealed no significant main effect of vocalizers' actual weights on listeners' weight assessments following Bonferroni correction ($b = 0.07, p = .016$). We found no effect of speech type on listener's assessments of weight ($b = -0.63$, CIs: --2.12, 0.85, $p = .404$), and no interaction between speech type and vocalizer actual weight ($b = 0.03$, CIs: -0.07, 0.006, $p = .099$). In male vocalizers, the effect of speech type was also not significant ($b = -1.63$, CIs: -3.178, -0.075, $p = .04$), nor was the effect of participant's actual weight on their assessed weight ($p = .084$), with no interaction between these two variables ($b = -0.02$, CIs: -0.07, 0.03 $p = .413$). Thus, these results suggest that the type of speech stimulus did not influence listeners' weight judgements (see Table S2.4). Whether participants completed the experiment in the lab or online did not significantly predict their weight assessments of either sex ($p = .875$ and $p = .982$ for female and male vocalizers, respectively). The same is true for stimulus duration ($p = .336$ and $p = .04$ for female and male vocalizers, respectively) but in females this relationship was slightly below the threshold of statistical significance after Bonferroni correction.

Assessment of psychosocial traits

Subsequent models examined listeners' assessments of socially relevant traits of vocalizers including attractiveness, dominance, likeability, trustworthiness, masculinity/femininity, and health. While many of these traits can be framed as both psychological and biological to some degree, the key distinction for the purposes of this study is that these traits were not defined by any objective ground-truth measure. Thus, while we could test whether listeners' perceptions of these traits varied when judging free versus read speech, we could not test whether speech

1
2
3 type affected the accuracy of these assessments, as we could for biological traits for which we had objective measures (sex, age, and body size).
4
5 The statistical models were thus composed of the same set of predictors (speech type, vocalizer/rater age, rater sex, method and stimulus
6
7 duration), however we did not assess accuracy in listeners' assessments of these psychosocial traits.
8
9

10
11 Speech type generally predicted differences in listeners' ratings of psychosocial traits, however the directions of these effects depended on
12
13 the specific trait. Read speech elicited relatively higher ratings of attractiveness (vocalizer sex: females $b = 0.51$, CIs: 0.74, -0.06, $p < .001$;
14
15 males $b = 0.41$, CIs: 0.18, 0.626, $p < .001$; Table S2.5), dominance (females $b = 1.274$, CIs: 1.002, 1.55, $p < .001$; males $b = 0.66$, CIs: 0.41, 0.91,
16
17 $p < .001$; Table S2.6), trustworthiness (females $b = 0.90$, CIs: 0.63, 1.18, $p < .001$; males $b = 0.82$, CIs: 0.56, 1.08, $p < .001$ Table S2.8) and
18
19 health (females $b = 0.51$, CIs: 0.27, 0.75, $p < .001$; males $b = 0.57$, CIs: 0.33, 0.82, $p < .001$) than did free speech. Likability ratings did not vary
20
21 significantly in females ($b = 0.26$, CIs: 0.01, 0.51, $p = .042$) and males ($b = 0.08$, CIs: -0.14, 0.30, $p = .477$; Table S2.7. Masculinity-femininity
22
23 ratings also did not vary between speech types for either female ($b = -0.04$, CIs: -0.22, 0.13, $p = .63$) or male vocalizers ($b = -0.008$, CIs: -0.18,
24
25 0.16, $p = .928$; Table S2.9).
26
27
28
29
30

31 In all cases, whether participants completed the experiment in the lab or online did not affect the pattern of results. However, stimulus
32
33 duration was a significant predictor of trait assessment in the case of: attractiveness in females ($b = -0.03$, CIs: -0.04, -0.01, $p < .001$) and males
34
35 ($b = -0.03$, CIs: -0.04, -0.01, $p < .001$), dominance in females ($b = -0.06$, CIs: -0.08, -0.05, $p < .001$), likability in females ($b = -0.03$, CIs: -0.08, -
36
37 0.01, $p = .003$), trustworthiness in females ($b = -0.03$, CIs: -0.05, -0.01, $p < .001$), trustworthiness in males ($b = -0.03$, CIs: -0.05, -0.01, $p < .001$),
38
39
40
41
42
43
44
45
46

1
2
3 health in females ($b = -0.03$, CIs: $-0.05, -0.02$, $p < .001$) and health in males ($b = -0.03$, CIs: $-0.04, -0.01$, $p = .002$). In each case longer stimulus
4
5 was related to lower rating of a given trait. No effect of stimulus length was observed in case of dominance in males ($p = .014$), likability in
6
7 males ($p = .146$) and femininity-masculinity in both females ($p = .988$) and males ($p = .583$).
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

For Review Only

Table S2.1*Model with sex accuracy rating an outcome*

Female vocalizers						
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals		
				Lower	Upper	
Fixed effects						
Intercept	0.987	0.003	<.001	0.980	0.993	
Stimulus type (0-free, 1-read)	0.001	0.009	0.893	-0.017	0.020	
Vocalizer's properties						
Age	-0.000	0.000	0.542	-0.000	0.000	
Rater's properties						
Sex (0-F, 1-M)	-0.008	0.000	0.203	-0.020	0.004	
Age	0.000	0.007	0.886	-0.001	0.001	
Method (0-online, 1-lab)	-0.012	0.007	0.097	-0.026	0.002	
Stimulus duration	-0.000	0.001	0.817	-0.001	0.001	
Random effects						
	Variance		ICC			
Rater's ID	0.002		0.170			
Vocalizer's ID	0.000		0.017			
Residual	0.012					
N _{raters} =429, N _{vocalizers} =101, N _{observations} =4290						
Male vocalizers						
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals		
				Lower	Upper	
Fixed effects						
Intercept	0.994	0.002	<.001	0.989	0.998	

1						
2						
3	Stimulus type (0-free, 1-read)	0.002	0.006	0.766	-0.010	0.014
4	Vocalizer's properties					
5	Age	0.000	0.000	0.894	-0.000	0.000
6	Rater's properties					
7	Sex (0-F, 1-M)	-0.006	0.004	0.164	-0.015	0.002
8	Age	-0.000	0.000	0.459	-0.001	0.000
9						
10						
11	Method (0-online, 1-lab)	-0.009	0.005	0.100	-0.019	0.002
12	Stimulus duration	-0.000	0.000	0.955	-0.001	0.001
13						
14	Random effects					
15		Variance		ICC		
16	Rater's ID	0.001		0.146		
17	Vocalizer's ID	0.000		0.003		
18	Residual	0.006				
19	N _{raters} =429, N _{vocalizers} =96, N _{observations} =3536					
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

View Only

Table S2.2

Model with rated age as outcome

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	31.475	0.463	32.382	30.568	32.382
Stimulus type (0-free, 1-read)	-0.311	0.773	1.205	-1.826	1.205
Vocalizer's properties					
Age	0.689	0.033	0.753	0.626	0.753
Vocalizer's age*Stimulus Type	0.034	0.018	0.069	-0.001	0.069
Rater's properties					
Sex (0-F, 1-M)	-1.136	0.489	-0.176	-2.095	-0.176
Age	0.020	0.023	0.065	-0.024	0.065
Method (0-online, 1-lab)	0.568	0.585	1.715	-0.580	1.715
Stimulus duration	0.054	0.052	0.155	-0.048	0.155
Random effects					
	Variance		ICC		
Rater's ID	17.885		0.260		
Vocalizer's ID	15.387		0.232		
Residual	50.819				
N _{raters} =405, N _{vocalizers} =101, N _{observations} =4047					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	31.420	0.559	<.001	30.325	32.516
Stimulus type (0-free, 1-read)	-0.487	0.722	0.499	-1.902	0.927

Vocalizer's properties						
Age	0.551	0.044	<.001	0.466	0.637	
Vocalizer's age*Stimulus Type	0.014	0.021	0.494	-0.026	0.054	
Rater's properties						
Sex (0-F, 1-M)	-0.849	0.510	0.096	-1.848	0.150	
Age	0.061	0.024	0.010	0.015	0.108	
Method (0-online, 1-lab)	-0.422	0.610	0.489	-1.618	0.774	
Stimulus duration	0.035	0.051	0.494	-0.065	0.136	
Random effects						
	Variance		ICC			
Rater's ID	18.910		0.286			
Vocalizer's ID	23.411		0.331			
Residual	47.245					
N _{raters} =405, N _{vocalizers} =96, N _{observations} =3325						

For Review Only

Table S2.3*Model with rated height as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	165.589	0.319	< .001	164.964	166.214
Stimulus type (0-free, 1-read)	1.569	0.601	0.009	0.390	2.747
Vocalizer's properties					
Age	-0.029	0.016	0.068	-0.060	0.002
Height	0.095	0.037	0.012	0.022	0.168
Vocalizer's height*Stimulus type	0.088	0.038	0.020	0.014	0.162
Rater's properties					
Sex (0-F, 1-M)	0.590	0.546	0.281	-0.481	1.661
Age	-0.022	0.023	0.340	-0.067	0.023
Method (0-online, 1-lab)	1.034	0.664	0.120	-0.268	2.337
Stimulus duration	-0.090	0.040	0.024	-0.169	-0.012
Random effects					
	Variance		ICC		
Rater's ID	22.309		0.382		
Vocalizer's ID	2.331		0.061		
Residual	36.016				
N _{raters} =362, N _{Vocalizers} =101, N _{observations} =3597					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	175.631	0.435	< .001	174.779	176.482

1						
2						
3	Stimulus type (0-free, 1-read)	-0.499	0.660	0.450	-1.793	0.795
4	Vocalizer's properties					
5	Age	-0.032	0.022	0.140	-0.075	0.010
6	Height	0.206	0.042	<.001	0.124	0.289
7	Vocalizer's height*Stimulus type	0.025	0.043	0.565	-0.059	0.109
8	Rater's properties					
9	Sex (0-F, 1-M)	0.460	0.738	0.533	-0.986	1.907
10	Age	-0.084	0.031	0.007	-0.146	-0.023
11						
12						
13	Method (0-online, 1-lab)	1.368	0.899	0.129	-0.394	3.131
14	Stimulus duration	0.030	0.047	0.522	-0.062	0.122
15						
16						
17	Random effects		Variance		ICC	
18	Rater's ID		41.945		0.483	
19	Vocalizer's ID		4.145		0.085	
20	Residual		44.814			
21	N _{raters} =362, N _{vocalizers} =96, N _{observations} =3021					
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.4*Model with rated weight as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	63.456	0.422	< .001	62.628	64.284
Stimulus type (0-free, 1-read)	-0.634	0.758	0.403	-2.120	0.852
Vocalizer's properties					
Age	0.240	0.028	< .001	0.184	0.295
Weight	0.074	0.030	0.016	0.015	0.133
Vocalizer's weight*Stimulus type	0.033	0.020	0.099	-0.006	0.072
Rater's properties					
Sex (0-F, 1-M)	0.350	0.582	0.548	-0.790	1.490
Age	-0.079	0.028	0.005	-0.134	-0.024
Method (0-online, 1-lab)	-0.113	0.749	0.880	-1.581	1.354
Stimulus duration	0.046	0.051	0.366	-0.054	0.146
Random effects					
	Variance		ICC		
Rater's ID	24.351		0.321		
Vocalizer's ID	8.402		0.140		
Residual	51.487				
N _{raters} =365, N _{vocalizers} =101, N _{observations} =3633					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	77.248	0.515	< .001	76.238	78.257

1						
2						
3	Stimulus type (0-free, 1-read)	-1.626	0.792	0.040	-3.178	-0.075
4	Vocalizer's properties					
5	Age	0.151	0.029	< .001	0.094	0.209
6	Weight	0.047	0.027	0.084	-0.006	0.101
7	Vocalizer's weight*Stimulus type	-0.019	0.024	0.413	-0.066	0.027
8	Rater's properties					
9	Sex (0-F, 1-M)	-2.006	0.813	0.014	-3.599	-0.414
10	Age	-0.064	0.039	0.104	-0.141	0.013
11						
12						
13	Method (0-online, 1-lab)	-0.184	1.044	0.860	-2.229	1.862
14	Stimulus duration	0.116	0.057	0.040	0.005	0.227
15						
16	Random effects		Variance		ICC	
17	Rater's ID		49.817		0.446	
18	Vocalizer's ID		7.569		0.109	
19	Residual		61.802			
20						
21	$N_{\text{raters}}=365, N_{\text{vocalizers}}=95, N_{\text{observations}}=2962$					
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.5*Model with rated attractiveness as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.135	0.076	<.001	3.986	4.284
Stimulus type (0-free, 1-read)	0.508	0.121	<.001	0.271	0.744
Vocalizer's properties					
Age	-0.028	0.005	<.001	-0.038	-0.018
Rater's properties					
Sex (0-F, 1-M)	-0.001	0.094	0.991	-0.185	0.183
Age	0.008	0.004	0.044	0.000	0.016
Method (0-online, 1-lab)	-0.148	0.108	0.173	-0.360	0.064
Stimulus duration	-0.027	0.008	<.001	-0.043	-0.011
Random effects					
	Variance		ICC		
Rater's ID	0.758		0.346		
Vocalizer's ID	0.360		0.200		
Residual	1.436				
N _{raters} =435, N _{vocalizers} =101, N _{observations} =4330					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.836	0.079	<.001	3.682	3.990

1						
2						
3	Stimulus type (0-free, 1-read)	0.405	0.113	< .001	0.184	0.626
4	Vocalizer's properties					
5	Age	-0.013	0.005	0.022	-0.023	-0.002
6	Rater's properties					
7	Sex (0-F, 1-M)	-0.465	0.101	< .001	-0.662	-0.268
8	Age	0.002	0.004	0.583	-0.006	0.011
9						
10						
11	Method (0-online, 1-lab)	-0.126	0.116	0.279	-0.354	0.102
12	Stimulus duration	-0.027	0.008	< .001	-0.043	-0.011
13						
14	Random effects		Variance		ICC	
15	Rater's ID		0.869		0.388	
16	Vocalizer's ID		0.342		0.200	
17	Residual		1.373			
18						
19	N _{raters} =435, N _{vocalizers} =96, N _{observations} =3577					
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.6*Model with rated dominance as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.644	0.076	< .001	3.496	3.792
Stimulus type (0-free, 1-read)	1.274	0.139	< .001	1.002	1.546
Vocalizer's properties					
Age	0.004	0.005	0.438	-0.006	0.014
Rater's properties					
Sex (0-F, 1-M)	0.012	0.096	0.900	-0.175	0.199
Age	0.011	0.004	0.014	0.002	0.019
Method (0-online, 1-lab)	0.096	0.113	0.394	-0.125	0.318
Stimulus duration	-0.063	0.009	< .001	-0.081	-0.045
Random effects					
	Variance		ICC		
Rater's ID	0.715		0.290		
Vocalizer's ID	0.343		0.164		
Residual	1.753				
N _{raters} =415, N _{vocalizers} =101, N _{observations} =4130					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.703	0.086	< .001	3.535	3.871
Stimulus type (0-free, 1-read)	0.657	0.128	< .001	0.406	0.908
Vocalizer's properties					

1						
2						
3	Age	0.005	0.006	0.455	-0.008	0.017
4	Rater's properties					
5	Sex (0-F, 1-M)	0.097	0.099	0.324	-0.096	0.290
6	Age	0.005	0.004	0.236	-0.003	0.014
7						
8						
9	Method (0-online, 1-lab)	0.084	0.116	0.472	-0.144	0.312
10	Stimulus duration	-0.023	0.009	0.014	-0.041	-0.005
11						
12	Random effects		Variance		ICC	
13	Rater's ID		0.740		0.308	
14	Vocalizer's ID		0.466		0.219	
15	Residual		1.665			
16						
17	N _{raters} =415, N _{vocalizers} =96, N _{observations} =3460					
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.7*Model with rated likability as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.437	0.073	< .001	4.294	4.579
Stimulus type (0-free, 1-read)	0.260	0.128	0.042	0.010	0.510
Vocalizer's properties					
Age	-0.016	0.005	0.001	-0.025	-0.006
Rater's properties					
Sex (0-F, 1-M)	-0.123	0.092	0.181	-0.303	0.057
Age	0.006	0.004	0.192	-0.003	0.015
Method (0-online, 1-lab)	0.136	0.121	0.260	-0.100	0.372
Stimulus duration	-0.025	0.008	0.003	-0.042	-0.009
Random effects					
	Variance		ICC		
Rater's ID	0.707		0.314		
Vocalizer's ID	0.313		0.169		
Residual	1.542				
N _{raters} =430, N _{vocalizers} =101, N _{observations} =4271					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.273	0.067	< .001	4.142	4.404
Stimulus type (0-free, 1-read)	0.080	0.113	0.477	-0.141	0.301
Vocalizer's properties					

1						
2						
3	Age	-0.002	0.004	0.729	-0.010	0.007
4	Rater's properties					
5	Sex (0-F, 1-M)	-0.215	0.091	0.019	-0.394	-0.036
6	Age	0.001	0.004	0.866	-0.008	0.009
7						
8						
9	Method (0-online, 1-lab)	-0.035	0.120	0.769	-0.270	0.200
10	Stimulus duration	-0.012	0.008	0.146	-0.028	0.004
11						
12	Random effects		Variance		ICC	
13	Rater's ID		0.680		0.324	
14	Vocalizer's ID		0.215		0.132	
15	Residual		1.415			
16						
17	N _{raters} =430, N _{vocalizers} =96, N _{observations} =3568					
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.8*Model with rated trustworthiness as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.031	0.071	<.001	3.891	4.170
Stimulus type (0-free, 1-read)	0.904	0.140	<.001	0.630	1.178
Vocalizer's properties					
Age	-0.000	0.004	0.953	-0.009	0.008
Rater's properties					
Sex (0-F, 1-M)	-0.091	0.103	0.380	-0.293	0.112
Age	0.005	0.005	0.334	-0.005	0.014
Method (0-online, 1-lab)	0.027	0.126	0.827	-0.219	0.274
Stimulus duration	-0.033	0.009	<.001	-0.051	-0.014
Random effects					
	Variance		ICC		
Rater's ID	0.838		0.306		
Vocalizer's ID	0.229		0.107		
Residual	1.905				
N _{raters} =399, N _{vocalizers} =101, N _{observations} =3971					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	3.842	0.071	<.001	3.702	3.982
Stimulus type (0-free, 1-read)	0.820	0.131	<.001	0.563	1.076
Vocalizer's properties					
Age	0.005	0.005	0.272	-0.004	0.014

Rater's properties						
	Sex (0-F, 1-M)	-0.019	0.102	0.849	-0.219	0.181
	Age	0.003	0.005	0.589	-0.007	0.012
	Method (0-online, 1-lab)	0.173	0.124	0.166	-0.071	0.416
	Stimulus duration	-0.032	0.009	<.001	-0.051	-0.014
Random effects		Variance		ICC		
	Rater's ID	0.776		0.302		
	Vocalizer's ID	0.222		0.110		
	Residual	1.794				

$N_{\text{raters}}=399$, $N_{\text{vocalizers}}=96$, $N_{\text{observations}}=3239$

For Review Only

Table S2.9*Model with rated femininity-masculinity as outcome. Higher score indicated higher masculinity*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	2.228	0.061	<.001	2.109	2.348
Stimulus type (0-free, 1-read)	-0.043	0.089	0.630	-0.217	0.132
Vocalizer's properties					
Age	0.008	0.003	0.003	0.003	0.014
Rater's properties					
Sex (0-F, 1-M)	-0.167	0.105	0.114	-0.374	0.040
Age	-0.007	0.005	0.162	-0.016	0.003
Method (0-online, 1-lab)	0.140	0.128	0.273	-0.110	0.391
Stimulus duration	0.000	0.006	0.988	-0.011	0.012
Random effects					
	Variance		ICC		
Rater's ID	1.007		0.562		
Vocalizer's ID	0.098		0.111		
Residual	0.784				
N _{raters} =407, N _{vocalizers} =101, N _{observations} =4047					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.677	0.066	<.001	5.548	5.807
Stimulus type (0-free, 1-read)	-0.008	0.087	0.928	-0.179	0.163
Vocalizer's properties					
Age	0.006	0.005	0.167	-0.003	0.015

Rater's properties					
Sex (0-F, 1-M)	-0.054	0.083	0.517	-0.217	0.109
Age	0.000	0.004	0.971	-0.007	0.008
Method (0-online, 1-lab)	-0.151	0.101	0.136	-0.349	0.047
Stimulus duration	0.003	0.006	0.583	-0.009	0.016

Random effects	Variance	ICC
Rater's ID	0.588	0.455
Vocalizer's ID	0.250	0.262
Residual	0.705	

$N_{\text{raters}}=407$, $N_{\text{vocalizers}}=96$, $N_{\text{observations}}=3375$

For Review Only

Table S2.10*Model with rated health as outcome*

Female vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	5.187	0.071	<.001	5.048	5.326
Stimulus type (0-free, 1-read)	0.510	0.122	<.001	0.272	0.749
Vocalizer's properties					
Age	-0.018	0.005	<.001	-0.027	-0.009
Rater's properties					
Sex (0-F, 1-M)	-0.040	0.091	0.661	-0.219	0.139
Age	0.005	0.004	0.200	-0.003	0.014
Method (0-online, 1-lab)	-0.038	0.112	0.735	-0.258	0.182
Stimulus duration	-0.032	0.008	<.001	-0.048	-0.016
Random effects					
	Variance		ICC		
Rater's ID	0.726		0.333		
Vocalizer's ID	0.294		0.168		
Residual	1.456				
N _{raters} =441, N _{vocalizers} =101, N _{observations} =4387					
Male vocalizers					
Parameter	Estimate	SE	<i>p</i>	95% confidence intervals	
				Lower	Upper
Fixed effects					
Intercept	4.975	0.080	<.001	4.819	5.131
Stimulus type (0-free, 1-read)	0.574	0.124	<.001	0.332	0.817

1						
2						
3	Vocalizer's properties					
4	Age	-0.010	0.006	0.094	-0.021	0.002
5	Rater's properties					
6	Sex (0-F, 1-M)	0.029	0.099	0.771	-0.165	0.222
7	Age	0.001	0.005	0.853	-0.008	0.010
8						
9						
10	Method (0-online, 1-lab)	-0.175	0.121	0.149	-0.413	0.062
11	Stimulus duration	-0.027	0.009	0.002	-0.044	-0.010
12						
13	Random effects			Variance		ICC
14	Rater's ID			0.824		0.354
15	Vocalizer's ID			0.365		0.196
16	Residual			1.502		
17						
18	N _{raters} =441, N _{vocalizers} =96, N _{observations} =3580					
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

Table S2.11

Actual self-reported properties of the Vocalizers

	Age (years)			Height (cm)			Weight (kg)		
	M (SE)	Min	Max	M (SE)	Min	Max	M(SE)	Min	Max
Females	35.39 (12.68)	19	67	165.65 (5.36)	152	176	65.00 (11.62)	45	100
Males	30.54 (11.80)	18	65	180 (6.18)	161	197	80.00 (12.95)	58	130
Total	33.03 (12.46)	18	67	172.70 (9.30)	152	197	72.99 (14.41)	45	130

For Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

Table S2.12.*Properties of the vocalizers rated by raters depending on stimulus type.*

Estimated	Sex accuracy	Age	Height (kg)	Weight (kg)	Attractiveness	Dominance	Likability	Trustworthiness	Femininity- Masculinity	Health
Female vocalizers										
Free speech	99% (0.005)	32.04 (0.54)	164.81 (0.44)	63.77 (0.57)	3.88 (0.10)	3.01 (0.10)	4.31 (0.10)	3.58 (0.10)	2.25 (0.08)	4.93 (0.09)
Read speech	99% (0.006)	30.91 (0.51)	166.37 (0.44)	63.14 (0.57)	4.39 (0.10)	4.28 (0.10)	4.57 (0.10)	4.48 (0.10)	2.21 (0.08)	5.44 (0.09)
Male vocalizers										
Free speech	99% (0.004)	31.66 (0.68)	175.88 (0.56)	78.06 (0.67)	3.63 (0.10)	3.38 (0.11)	4.23 (0.09)	3.43 (0.10)	5.68 (0.08)	4.69 (0.10)
Read speech	99% (0.004)	31.18 (0.65)	175.38 (0.53)	76.43 (0.63)	3.04 (0.09)	4.03 (0.10)	4.31 (0.08)	4.25 (0.09)	5.67 (0.08)	5.26 (0.10)

*Note. Estimated marginal means are calculated with control for all covariates from the models. Standard errors are presented in parentheses. The values that differed significantly are marked as **bolded**.*