Do voices carry valid information about a speaker’s personality?

Julia Stern (Jünger)¹, Christoph Schild², Ruben Arslan³, Benedict C. Jones⁴, Lisa M. DeBruine⁴, Amanda Hahn⁵, David A. Puts⁶, Ingo Zettler², Tobias L. Kordsmeyer¹, David Feinberg⁷, Dan Zamfir⁸, & Lars Penke¹

¹Department of Psychology & Leibniz ScienceCampus Primate Cognition 
University of Goettingen 
Gosslerstrasse 14, 37073 Goettingen, Germany

²Department of Psychology 
University of Copenhagen 
Øster Farimagsgade 2A, 1353 Copenhagen, Denmark

³Max Planck Institute for Human Development 
Lentzeallee 94, 14195 Berlin, Germany

⁴Institute of Neuroscience & Psychology 
University of Glasgow 
Hillhead Street, Glasgow, Scotland, UK

⁵Department of Psychology 
Humboldt State University 
1 Harpst Street, Arcata, CA 95521, USA

⁶Department of Anthropology & Center for Brain, Behavior and Cognition 
Pennsylvania State University 
University Park, PA 16802, USA

⁷Department of Psychology, Neuroscience, and Behaviour 
McMaster University 
Hamilton, Ontario L8S 4K1, Canada

⁸Developmental Psychology and Education 
University of Toronto 
63 St. George Street Toronto, ON Canada M5S

Corresponding author: Julia Stern (julia.juenger@psych.uni-goettingen.de)
VOICES AND SPEAKER’S PERSONALITY

Abstract

Research on links between peoples’ personality traits and their voices has primarily focused on other peoples’ personality judgments about a target person based on a target person’s vocal characteristics, particularly voice pitch. However, it remains unclear whether individual differences in voices are linked to actual individual differences in personality traits, and thus whether vocal characteristics are indeed valid cues to personality. Here, we investigate how the personality traits of the Five Factor Model of Personality, sociosexuality, and dominance are related to fundamental frequency (voice pitch) and formant frequencies. For this purpose, we will conduct a secondary data analysis of a large sample (2,133 participants) from eleven different, independent datasets with a Bayesian approach.
VOICES AND SPEAKER’S PERSONALITY

Theoretical background

When meeting new people, we make spontaneous inferences and form first impressions about a wide range of characteristics (e.g. Ambady & Rosenthal, 1992). Besides physical characteristics, such as sex (Bachorowski & Owren, 1999; Puts et al., 2012), body size (Feinberg et al., 2005; Xu et al., 2013), or strength (Sell et al., 2010), we also form stable impressions about other relevant characteristics, including attitudes, intentions, values, beliefs and personality traits (Borkenau & Liebler, 1992; Borkenau et al., 2004; McAleer et al., 2014; Scherer, 1972).

While other peoples’ physical appearance might be an important cue to our social evaluations (Naumann et al., 2009), peoples’ voices are another factor that influences socially relevant impressions (Borkenau & Liebler, 1992; Mileva et al., 2018; Zuckerman & Driver, 1989). Indeed, when visual cues are absent, e.g., when listening to the radio or hearing a voice on the telephone, people still form judgments about others, based only on acoustic information (Borkenau & Liebler, 1992; Mileva et al., 2018). Human voices and judgments based on their sound seem to have an association with important life outcomes: Studies have reported that voice characteristics predict mate choice (for an overview see Puts et al., 2014), courtship outcomes (Leongómez et al., 2014), and reproductive success (Apicella et al., 2007). Even voting behavior has been reported to be influenced by politicians’ voices, in that participants preferred to vote for candidates with a lower voice pitch (the rate of vocal fold vibrations which influences perceptions of pitch, usually equated with fundamental frequency), presumably because low pitch sounds more dominant, honest, intelligent, and attractive (Klofstad et al., 2016; Tigue et al., 2012). Further, CEOs with lower voice pitch oversee larger companies, receive higher compensations, and enjoy longer tenures than CEOs with higher voice pitch (Mayew, Parsons, & Venkatachalap, 2013).
VOICE pitch has been associated with personality judgments in multiple studies, showing that men and women with lower voice pitch are perceived as more dominant (e.g. Borkowska & Pawlowski, 2011; Collins, 2000; Hodges-Simeon et al., 2010; Puts et al., 2006; 2007). Furthermore, people with higher pitched voices have been reported to be more nervous (Apple et al., 1979), less agreeable (Scherer, 1978), and higher in neuroticism (Aronovitch, 1976; Scherer, 1978). Moreover, men with lower voice pitch and lower formant frequencies (resonant frequencies determined by the length and shape of the vocal tract and influence perceptions of vocal timbre) are perceived as more attractive (e.g. Collins, 2000; Feinberg et al., 2011; Hodges-Simeon et al., 2010; Jünger et al., 2018b; Puts 2005; 2006), while vocal attractiveness correlates positively with rated conscientiousness and negatively with rated neuroticism (Zuckerman et al., 1995). These social evaluations and personality judgments based on other peoples’ voices are characterized by a high level of agreement between perceivers across different speech contents and contexts (Mahrholz et al., 2018; McAleer et al., 2014; Scherer, 1972). Interestingly, first studies have indicated that personality judgments based on voice is, indeed, somewhat accurate (compared with target people’s personality self-ratings), at least for extraversion and emotional stability (Borkenau & Liebler, 1992, with N = 100; Scherer, 1972, with N = 59; Scherer, 1978, with N = 24).

While there are some studies on personality judgments from voices, literature on vocal characteristics and their actual link to target personality and individual differences is rather scarce. Only one early study has reported direct associations between some vocal characteristics and personality trait variables: Allport and Cantril (1934) reported that more extraverted people had ‘louder, more boisterous and carefree voices’ (in N = 3 speakers scoring low, medium and high on extraversion, respectively). Moreover, a relationship between lower voice pitch and markers associated with more unrestricted sociosexual behavior has been reported in that lower voice pitch in men, as well as more attractive voices.
in both sexes, were associated with self-reporting a higher number of sex partners (Hughes et al., 2004; Puts, 2005).

Nevertheless, no study has directly investigated links between measured acoustic parameters and personality traits. Therefore, this study’s aim is to examine which vocal characteristics, if any, are related to self-reported personality traits. For this purpose, we will combine different independent datasets from previous studies for secondary data analysis, resulting in a large sample size to investigate the relationship between vocal characteristics and personality traits.

**Hypotheses**

Based on previous studies, we hypothesize that voice pitch is a valid cue to the speaker’s personality:

Hypothesis 1: Participants with lower voice pitch will have a more dominant personality.

Hypothesis 2: Participants with lower voice pitch will score higher on agreeableness.

Hypothesis 3: Participants with lower voice pitch will score lower on neuroticism.

Hypothesis 4: Participants with lower voice pitch will report having a more unrestricted sociosexuality.

We will perform a number of exploratory analyses investigating the relationships between voice pitch and conscientiousness, extraversion and openness for experiences. Additionally, we will investigate relationships between personality traits and the acoustic parameter formant position ($P_f$; Puts et al., 2012). If we find support for Hypothesis 4, we will also investigate the associations of voice parameters with the three facets of the SOI-R (Penke & Asendorpf, 2008).
VOICES AND SPEAKER’S PERSONALITY

Methods

Participants
A total of 2,133 participants (n = 931 men; n = 1,202 women; aged 16 to 56 years) were recruited in eleven different, independent previous studies focused on other research questions (see Tables 1 and 2 for more information).

Measures

Personality variables
All personality measures were taken via self-report questionnaires: Sociosexuality was measured with the SOI-R (Penke & Asendorpf, 2008), dominance with the Interpersonal Adjective List (IAL; Jacobs & Scholl, 2005) or the German version of the revised Interpersonal Adjective Scale (IAS-R; Ostendorf, 2001). Neuroticism, extraversion, openness to experience, agreeableness and conscientiousness were assessed as the dimensions of the Five Factor Model (FFM) of Personality and measured with the German NEO-FFI (Borkenau & Ostendorf, 1993; 12 items per dimension; Sample 2), the BFI-10 (Rammstedt & John, 2007; Sample 9), the BFI 42-item version (Lang et al., 2001; Samples 3 and 6), or the BFI 44-item version (John, Donahue, & Kentle, 1991; Samples 7, 8 and 10). Detailed information are shown in Table 1.

Voice recordings and analyses
For voice recordings, participants were instructed to either read an excerpt from a standardized voice passage (e.g. the “rainbow passage”; Fairbanks, 1960), count from 1 to 10, say “A-E-I-O-U” (speaking vowels), say a standardized sentence (“Hi, I am a student at McMaster University”), or present themselves (“What do you think is great about yourself?”). Detailed information on voice recordings used in the different datasets are shown in Table 1. Length and content of different voice recordings should not affect relationships between personality and vocal parameters, because vocal parameters usually show moderate to strong correlations across different recordings, even independent of length and content (Mahrholz et
VOICES AND SPEAKER’S PERSONALITY

al., 2018, Puts et al., 2012). Moreover, all recordings are of a neutral content, in which pitch variation is usually very small (Belin et al., 2008). For all samples with multiple voice recordings per participant (due to a within-subjects design with repeated measures), we will analyze the recordings from the first session only.

All voices will be analyzed using Praat software (Version 6.0.37; Boersma & Weenink, 2018). We will measure the following variables: mean $F_0$ (fundamental frequency), mean formant frequencies (supralaryngeal vocal tract resonances) from which we will compute $P_f$ (by standardizing $f_1$, $f_2$, $f_3$ and $f_4$ and dividing the sum by four, following Puts et al., 2012).

Scripts for acoustic analyses are publicly available on the OSF (Feinberg, 2018; Puts & Cardenas, 2018).

Data transformations

For $F_0$ analyses, we will use the variable “mean Pitch”, extracted by Praat. $P_f$ will be computed using the standardized and aggregated value of the four formants (Puts et al., 2012).

All personality measures will be z-standardized.

Statistical analyses

All analyses will be computed with the statistical software R (R Core Team, 2016) and the package brms (Bürkner, 2017) which implements an R interface to the probabilistic programming language Stan (Carpenter et al., 2015). The analysis code is publicly available¹ (https://osf.io/x4jzq/?view_only=936b51fe701b4fc68ce9ece565f6292a). According to recommendations by Kruschke (2018) we will focus on estimating the strength of associations between voice parameters and personality traits. However, to give recommendations for future research, we will implement a decision rule (Makowski & Lüdecke, 2019) where we combine a region of practical equivalence (ROPE) from -0.1 to 0.1 with the 95% highest

¹ The analysis code has been produced using simulated data. It currently only includes analyses for the relationship between vocal characteristics and dominance, but analyses for all other personality traits will be identical, but with the respective trait as outcome.
density interval (HDI) of the estimated effect sizes. This will allow us to differentiate between three scenarios: a) The estimated HDI is completely within the ROPE. Future researchers should not expect to find substantive associations here, unless they think our ROPE was too broad, or can substantially improve on our measurement of voice parameters or dependent variables. b) The HDI overlaps with the ROPE, so we do not know whether the association is outside the ROPE – in other words, our estimates are insufficiently precise and future research with larger samples is needed. c) The estimated effect sizes seem substantial because our HDI is outside the ROPE. Future research should work to better characterize these associations.

As all variables except age and sex will be standardized, we will use weakly informative normal priors (centered on zero, with standard deviations of 3) for all population-level effects and the brms default priors for all other parameters. All models will be adjusted for age and sex. Sex will be effect-coded without weights, with women coded as -1 and men as 1. Age will be adjusted as a linear effect. For one sample, we lack precise age information; it was only recorded that participants were undergraduates. Therefore, we will use an errors-in-variables approach. In datasets where age was measured, age is entered as measured with a standard error of 0.5 (rounding error). In the undergraduate dataset with unmeasured ages, we assume an age of 20 with a standard error of 3. In effect, this means this dataset will not contribute much to estimating the age effect, but does not have to exclude because of missing data. Because the personality traits were measured with scales of varying length, we will conduct a robustness check in which we allow not only the intercepts and slopes to vary by study, but also the residual variation.

We will always fit one model per personality trait to be predicted. In Wilkinson notation, the model will be specified as:

Personality_outcome ~ F_0 + P_f + sex + me(age, age_se) + (1 | dataset)
To diagnose nonlinearity, we will graph the bivariate relationships between all vocal parameters and all traits in scatterplots overlaid with thin-plate spline smooths (Wood et al. 2016). If visual diagnosis indicates nonlinearity or interactions with sex for certain parameter-outcome combinations, we will fit models allowing nonlinearity via thin-plate splines and/or interactions, respectively. If approximative leave-one-out cross-validation (LOO-IC; Vehtari, Gelman, & Gabry, 2016) favors these adapted models over the simple main effect model (LOO-IC lower than by more than 2 standard errors), we will discuss these models instead. We will not apply the HDI+ROPE decision rule to nonlinear effects, but simply show them visually and discuss them. If we find that the HDI is not within the ROPE or if we find evidence for nonlinear effects, we will also fit an additional model to see whether this association is invariant across datasets by allowing the relevant linear terms in the regression to vary by dataset and comparing models’ LOO-ICs.
Extended Abstract for Registered Report in JRP special issue

Table 1

Information about the combined datasets on sample sizes, content of voice recordings, used personality questionnaires and publications in which the sample has been used before

<table>
<thead>
<tr>
<th>Dataset number</th>
<th>N  (male/female)</th>
<th>Voice recordings</th>
<th>SOI-R</th>
<th>BFI</th>
<th>NEO-FFI</th>
<th>IAL/IAS-R</th>
<th>Sample characteristics detailed in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400 (0/400)</td>
<td>Vowels</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Jones et al. (2018)</td>
</tr>
<tr>
<td>2</td>
<td>382 (190/192)</td>
<td>Counting 1-10</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>Asendorpf, Penke &amp; Back (2011)</td>
</tr>
<tr>
<td>3</td>
<td>284 (141/143)</td>
<td>Counting 1-10</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Penke &amp; Asendorpf (2008 Study 2)</td>
</tr>
<tr>
<td>4</td>
<td>187 (61/126)</td>
<td>Rainbow passage</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Puts et al. (2016)</td>
</tr>
<tr>
<td>5</td>
<td>186 (186/0)</td>
<td>German rainbow passage</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Schild, Stern, &amp; Zettler (2019)</td>
</tr>
<tr>
<td>6</td>
<td>165 (165/0)</td>
<td>Self- presentation</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Kordsmeyer &amp; Penke (2019)</td>
</tr>
<tr>
<td>7</td>
<td>157 (0/157)</td>
<td>German rainbow passage</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Jünger et al. (2018a; 2018b Study 2)</td>
</tr>
<tr>
<td>8</td>
<td>120 (0/120)</td>
<td>German rainbow passage</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>108 (44/64)</td>
<td>Standardized sentence</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>88 (88/0)</td>
<td>Vowels</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>56 (56/0)</td>
<td>Rainbow passage</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Hill et al. (2013)</td>
</tr>
</tbody>
</table>
Extended Abstract for Registered Report in JRP special issue

Table 2
Additional information about the combined datasets on age span and dropouts

<table>
<thead>
<tr>
<th>Dataset number</th>
<th>N (male/female)</th>
<th>Age span</th>
<th>Dropouts</th>
<th>Reasons for dropouts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400 (0/400)</td>
<td>16-30 years</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>382 (190/192)</td>
<td>18-54 years</td>
<td>(n = 1) for SOI-R and dominance analyses</td>
<td>Did not fill out these questionnaires</td>
</tr>
<tr>
<td>3</td>
<td>284 (141/143)</td>
<td>19-30 years</td>
<td>(n = 1) for dominance analyses</td>
<td>Technical problems</td>
</tr>
<tr>
<td>4</td>
<td>187 (61/126)</td>
<td>18-27 years</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>186 (186/0)</td>
<td>18-56 years</td>
<td>(n = 2) dropped out</td>
<td>Did not fill out the questionnaires</td>
</tr>
<tr>
<td>6</td>
<td>165 (165/0)</td>
<td>18-34 years</td>
<td>(n = 1) dropped out for (F0) analyses, (n = 5) for (Pf) analyses</td>
<td>Did not want their recording to be used for further analyses ((n = 1)) or technical problems ((n = 4))</td>
</tr>
<tr>
<td>7</td>
<td>157 (0/157)</td>
<td>18-35 years</td>
<td>(n = 15) dropped out for analyses including the BFI and SOI-R</td>
<td>Did not fill out these questionnaires</td>
</tr>
<tr>
<td>8</td>
<td>120 (0/120)</td>
<td>18-35 years</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>108 (44/64)</td>
<td>Undergraduates at McMaster University</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>88 (88/0)</td>
<td>19-31 years</td>
<td>none</td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>56 (56/0)</td>
<td>18-23 years</td>
<td>(n = 9) dropped out</td>
<td>Did not fill out the questionnaire or technical problems</td>
</tr>
</tbody>
</table>
References


Extended Abstract for Registered Report in JRP special issue


Extended Abstract for Registered Report in JRP special issue


Extended Abstract for Registered Report in JRP special issue


