2023-02-06

Trait impressions from voices are formed rapidly within 400ms of exposure

Mileva, Mila

10.1037/xge0001325
Journal of Experimental Psychology: General
American Psychological Association

Trait impressions from voices are formed rapidly within 400ms of exposure

Mila Mileva[1*] & Nadine Lavan[2*]

[*] The authors contributed equally

[1] School of Psychology, University of Plymouth, UK
[2] Department of Biological and Experimental Psychology, Queen Mary University of London, UK

Correspondence to:
Mila Mileva, School of Psychology, Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, United Kingdom. E-mail: mila.mileva@plymouth.ac.uk

or

Nadine Lavan, Department of Biological and Experimental Psychology, School of Behavioural and Biological Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, United Kingdom. E-mail: n.lavan@qmul.ac.uk

Word Count: 8423

Acknowledgements
The research leading to these results has received funding from a British Academy Postdoctoral Fellowship (PF20\100034) awarded to Mila Mileva and a Sir Henry Wellcome Fellowship (220448/Z/20/Z) awarded to Nadine Lavan.

Author Note
The data underlying the reported analyses have been uploaded to the OSF: https://osf.io/zjwcq/

Abstract

When seeing a face or hearing a voice, perceivers readily form first impressions of a person's characteristics – are they trustworthy, do they seem aggressive? One of the key claims about trait impressions from faces and voices alike is that these impressions are formed rapidly. For faces, studies have systematically mapped this rapid time course of trait impressions, finding that they are well-formed and stable after approximately 100ms of exposure. For voices, however, no systematic investigation of the time course of trait perception exists. In the current study, listeners provided trait judgements (attractiveness, dominance, trustworthiness) based on recordings of 100 voices that lasted either 50ms, 100ms, 200ms, 400ms, or 800ms. Based on measures of intra- and inter-rater agreement as well as correlations of mean ratings for different exposure conditions, we find that trait perception from voices is indeed rapid. Unlike faces, however, trait impressions from voices require longer exposure to develop and stabilise although they are still formed by 400ms. Furthermore, differences in the time course of trait perception from voices emerge across traits and voice gender: The formation of impressions of attractiveness and dominance required less exposure when based on male, rather than female, voices, whereas impressions of trustworthiness evolved over a more gradual time course for male and female voices alike. These findings not only provide the first estimate of the time course of voice trait impressions, but they also have implications for voice perception models where voices are regarded as "auditory faces".

Keywords: First impressions; time course; voice; trustworthiness; dominance; attractiveness; gating task

Introduction

Human faces and voices are both rich sources of person information in our social world – they can, very efficiently, provide us with impressions of the people we interact with such as their age, gender, body size, health condition, emotional state, identity, and even their personality (Bruce & Young, 1986; Schweinberger et al., 2014). Thus, despite being signals from different modalities, with physical properties that cannot be readily compared to one another, many parallels have been drawn between face and voice processing over the last 20 years. These parallels range from the existence of face- and voice-selective areas in the brain (Kanwisher & Yovel, 2006; Leaver & Rauschecker, 2010) and impairments specific to face and voice identity recognition (Neuner & Schweinberger, 2000) to similarities in the functional demands of face and voice recognition such as the need to separate stable identity signals from the substantial natural variability present during each separate encounter with the same person (Jenkins et al., 2011; Lavan et al., 2019). In fact, the most influential model of voice processing is heavily based on earlier face perception models, suggesting almost entirely comparable processing stages, and describing the voice as an 'auditory face' (Belin et al., 2011).

Another aspect of face and voice perception for which many parallels have been identified is trait perception. Trait perception describes the process where upon hearing a voice or seeing a face, perceivers readily form a wealth of impressions of what they believe someone might be like as a person: Are they likeable? Aggressive? Outgoing? These so-called trait impressions have been shown to affect our decisions and behaviours in relation to other people in a number of real-life situations, such as who we vote for in an election (Klofstad, 2016; Mileva et al., 2020; Sussman et al., 2013; Tigue et al., 2012), which AirBnB host we choose (Ert et al., 2016), and how we sentence people as part of a court case (Chen, Halberstam, & Yu, 2016; Wilson & Rule, 2015). For voices, these consequences of forming a positive or negative evaluation of people have been linked to both low-level acoustic aspects of the sound of someone's voice (e.g., F0; Klofstad et al., 2016; Tigue et al., 2012) as well as

aspects of (socio-)linguistic or phonetic cues, such as a person's accent (Bayard et al., 2001, Purnell et al., 1999). Trait impressions based on faces and voices similarly share a 2- or 3-dimensional structure with trustworthiness, dominance, and attractiveness as the main underlying dimensions (McAleer et al., 2014; Oosterhof & Todorov, 2008; Sutherland et al., 2013).

When investigating trait impressions from faces and voices alike, two features are usually highlighted in studies: Trait impressions 1) are shared across perceivers and 2) are formed rapidly. This shared nature is often quantified via measures of inter-rater agreement, such as Cronbach's alpha. Agreement measures show that e.g., a voice that is perceived to be highly trustworthy by one perceiver is also likely to be perceived as overall trustworthy by other perceivers (Lavan et al., 2021; Lavan, Mileva et al., 2021; Mahrholz et al., 2019; McAleer et al., 2014; Mileva et al., 2020, see also Kramer et al., 2018 for faces). While this consistency of trait impressions across perceivers points to some shared basis of these impressions, this does not necessarily mean that perceivers are able to glean insights into the true characteristics of a person. Instead, trait impressions are thought to primarily be reflections of shared stereotypes inferred from the physical properties of faces and voices (Klofstad & Anderson, 2018; Todorov, 2017; see also Lavan, Mileva et al. 2021).

The rapid formation of trait impressions is usually demonstrated via the presence of good inter-rater agreement. In the absence of being able to assess the objective accuracy of trait impressions, researchers therefore take good inter-rater agreement as evidence that perceivers' responses are not random but that they reflect meaningful percepts. Initial studies of trait perception in faces and voices observed high inter-rater agreement after relatively brief stimulus exposure (e.g., Todorov et al., 2005; McAleer et al., 2014). Consequently, studies of face perception then attempted to identify the minimum exposure to a face that is required to form a meaningful trait impression while also tracking how these impressions evolve over time.

Willis and Todorov (2006) presented participants with faces for 100ms, 500ms, and 1000ms. The authors then correlated the mean ratings from each of the different exposure durations with mean ratings from a task where participants provided trait ratings in a self-paced manner, such that there were no experimenter-driven time constraints. Correlations were significant for all three time-constrained conditions with no significant differences in correlation strength occurring. Similarly, the amount of variance in the time-unconstrained ratings that was explained by each of the time-constrained ratings was high (> 70%) but did not increase with additional exposure. Based on the correlations, the authors thus conclude that trait impressions from faces are already well-formed and stable after 100ms of exposure, since they resemble time-unconstrained trait impressions. Todorov et al. (2009) built on the findings of Willis and Todorov (2006) by including additional, shorter exposure durations (17ms, 33ms, 50ms, 67ms) for impressions of trustworthiness. The authors report weak but significant correlations with the time-unconstrained trait ratings after 33ms (r = .22), with correlation strength increasing up to 100ms of exposure, after which correlations stabilised. No significant correlations were found for the shortest exposure duration of 17ms. This study therefore suggests that trait impressions of trustworthiness can be formed from as little as 33ms, although around 100ms of exposure to a face are required for trait impressions to stabilise. South Palomares and Young (2018) replicate and extend these findings to trustworthiness, attractiveness, and status impressions, reporting both significant inter-rater agreement and significant correlations with time-unconstrained ratings for trait impressions based on a 33ms exposure. Similar findings were also reported by Bar et al. (2006) for the perception of threat from faces, where significant correlations between ratings based on 39ms of exposure were significantly correlated with ratings provided after 1700ms, while no significant correlation emerged after 26ms of exposure. Bar et al. (2006) furthermore make the point that the time course of trait impressions is, however, likely to differ to some degree for different trait impressions, as 39ms of exposure was not sufficient to assess the intelligence of a person based on their face.

To our knowledge, for voices, the shortest stimulus durations reported in the literature in the presence of good inter-rater agreement for trait impressions used recordings of a single word ("Hello" or "Hola") that lasted 300ms to 400ms (Baus et al., 2019; Mahrholz et al., 2018; McAleer et al., 2014; see also Purnell et al., 1999 for accent perception from a single word). While these stimuli are certainly short, justifying the claim that trait impressions from voices are formed rapidly, the studies did not specifically aim to establish the time course of voice impression formation. As a result, it is still unclear how quickly trait impressions are formed and how they develop over the course of the initial hundreds of milliseconds of exposure to a voice.

Based on the strict interpretation of the 'auditory face' metaphor and all already established similarities in face and voice trait impressions, one prediction could be that the time course of the development of face and voice trait impressions is very similar (i.e., stable voice impressions formed in approximately 100ms). However, recent work highlights important functional and neurological differences in face and voice perception (Schirmer, 2018; Young et al., 2020), while there are also numerous physical differences between faces and voices (e.g., the usually static nature of face stimuli vs. the inherently dynamic nature of voices). It therefore seems more likely that there might at least be some differences in how quickly trait impressions are established for faces vs voices.

In the current study, we therefore set out to systematically track the time course over which trait impressions from voices can be established. Specifically, we aimed to identify 1) what the minimal exposure duration is for shared voice trait impressions to emerge and 2) trace when trait impressions reach a degree of stability. Following the studies of Willis and Todorov (2006) and Todorov et al. (2009), we implemented a gating experiment for voice impressions of attractiveness, dominance, and trustworthiness. These three types of trait impressions were chosen since they have been shown to form the primary axes of vocal trait perception spaces (e.g., McAleer et al., 2014). In the gating experiment, participants provided trait ratings for 100

voices (50 female, 50 male) based on recordings of the vowel /a/ that were presented across a range of exposure duration, or gates: 50ms, 100ms, 200ms, 400ms and 800ms. This type of stimuli was chosen because sustained vowels are quasi-periodic in nature, i.e., evolve minimally over time. Vowels also include no substantial meaningful linguistic content, such that trait impressions are primarily based on the sound of the voice (see Methods). Based on the findings from the face and voice perception literatures (Baus et a. 2019; Mahrholz et al., 2018; McAleer et al., 2014), we predict that trait impressions for voices should be well-formed and relatively stable after 400ms of exposure. If trait perception from voices follows the time course of trait impressions from faces, we would furthermore expect that trait impressions are already well-established after 100ms of exposure (Willis & Todorov, 2006; Todorov et al.,2009). Following Bar et al. (2006), we would finally expect that different trait impressions might require different amounts of exposure to become well-formed and stable.

Methods

Participants

All participants were recruited via the local participant pools at Queen Mary University of London and the University of Plymouth. The study was approved by the local ethics committees. Participants received course credits after completing the experiment. We recruited a total of 289 participants. From this full sample, 134 participants were excluded due to having missed more than 20% of vigilance trials (see Procedure). Additionally, 4 participants were excluded due to giving the same rating/response across all voices for over 90% of trials in one or more of the three trait rating scales completed in the task.

We analysed the data from 151 participants (mean age = 21.2, SD = 6; 124 female, 1 non-binary). Of these 151 participants, 28 participants provided trait ratings for voice samples presented for 50ms, 30 participants for voice samples of a duration of 100ms, 27 participants for voice samples of a duration of 200ms, 34 participants for voice samples of a duration of

400ms, and 32 participants for voice samples of a duration of 800ms. The number of participants in each timing condition fits within the suggested sample size needed in order to achieve a stable mean trait impression rating from faces (Hehman et al., 2018, 26 raters for trustworthiness, 29 raters for attractiveness and 31 raters for dominance) and is also consistent with previous voice impression studies (e.g., ratings from 32 participants in McAleer et al., 2014).

Materials

Recordings of 100 German talkers (50 female, 50 male) aged 20-22 years were selected from the Saarbrücker Stimmdatenbank (Pützer et al., n.d.). For each recording, speakers produced the German vowel /a/ in a sustained manner, lasting > 800ms. All recordings included the steady-state portion of the vowels, such that e.g., onsets and offsets had been trimmed in most of these raw recordings.

From these 100 raw recordings, we then created the stimuli for the five duration conditions that were included in our experiment: Each of the raw stimuli were trimmed to a duration of 1) 50ms, 2) 100ms, 3) 200ms, 4) 400ms, and 5) 800ms. The starting points for these stimuli were always the same across exposure durations, such that e.g., the first 50ms of the stimuli trimmed to 100ms were identical to the stimuli trimmed to 50ms and so on. 25ms of fade-in and fade-out were applied to the beginning and end of the trimmed recordings respectively to improve the perceptual experience of the sounds. 50ms was chosen as the shortest duration as very short recordings of vowels are not at all times recognisable as being voices (instead of just generic bursts of sound). The longest duration of 800ms was chosen as a reflection of the available materials and because McAleer et al. (2014) report high inter-rater agreement for stimuli with an average duration of ~390ms, suggesting that trait impressions have been reliably formed following that amount of exposure. We note that using 25ms fade-in and fade-out for all stimuli means that there is no time at which stimuli trimmed to a duration of 50ms remain at full volume, while such a volume plateau exists for stimuli with a longer duration.

This lack of a volume plateau for stimuli trimmed to 50ms does, however, not seem to affect our findings: All measures associated with this shortest exposure condition behave in alignment with the trends observed for the longer exposure condition in almost all cases (see below).

The specific type of stimuli, sustained vowels, were chosen because the acoustic and thus voice-related information included across the entire recording is as quasi-steady-state or stable as possible for naturally produced voice stimuli. As such, any excerpt of the recording should 1) be largely representative of the full recording and should 2) be largely comparable to any other excerpt from the same recording, which is a prerequisite for perceptual gating studies. In other words, in our stimuli, an excerpt from close to the start of the recording of the vowel /a/ should be as similar as is possible in naturally produced stimuli in its acoustics to an excerpt from close to the end of the recording of the sustained vowel /a/. If we had, for example, used stimuli with linguistic content (e.g., words or sentences), the acoustic information would have evolved dynamically over time to include the different speech sounds in words. In this case, different samples would have included substantially different types of acoustic properties and would have thus been differentially informative for trait impressions (e.g., compare the acoustic properties of /h/ and /i/ in the word “hiss”), which would have confounded the gating experiment. Due to the nature of the stimuli, the current study thus examines trait impressions based on the sound of the voice as conveyed by the fundamental frequency (F0, or perceived pitch of the voice), formant frequencies, and periodicity of recordings.

In addition to these experimental stimuli, we furthermore created voice recordings via a text-to-speech synthesiser instructing participants to respond with a certain rating (“Please select 1 now”) to be used during the attention checks included in the main task (see also participant exclusions above).

Procedure

Testing was conducted online via the platform Qualtrics. Participants were presented with an information sheet, provided informed consent, and completed a check to ensure that audio playback was working adequately before starting the main experiment. This involved participants being presented with a stimulus, similar to the ones used throughout the experiment, that they could play multiple times in order to make sure they can hear the recording clearly and adjust their volume levels. In the main experiment, participants provided trait ratings on a scale from 1 to 9 for all 100 voices along the three underlying dimensions of trait impressions: attractiveness, dominance, and trustworthiness (McAleer et al., 2014; Oosterhof & Todorov, 2008; Sutherland et al., 2013). Participants were asked to provide these trait impressions based on their "gut feeling". They were also made aware that stimuli might be very brief, such that they were required to listen closely. During the task, participants could listen to the stimuli only once per trial and were asked to manually initiate playback in order to ensure they were ready to listen to each of the often very short recordings. For each rating scale, 1 indicated that the voice was not perceived to sound attractive/dominant/trustworthy at all, 9 indicated that the voice was perceived to sound extremely attractive/dominant/trustworthy.

Participants were randomly assigned to one of the five exposure conditions (50ms, 100ms, 200ms, 400ms, and 800ms) included in our experiment: This means that participants were exposed to stimuli of only one duration throughout the main experiment. All voices were rated once per trait, with the exception of 20 voices (10 female, 10 male), which were repeated twice per trait, enabling us to examine how reliable each participant's impressions were within each trait. We used a blocked design with each block containing ratings of the same trait and the order of the blocks was randomised across participants. Both male and female voices were presented within the same block. The order of the voices was fully randomised across blocks and participants. As such, a participant in the 100ms condition might therefore first rate all 100 voices (+ 20 repeated stimuli) based on recordings that were trimmed to 100ms for e.g.,

attractiveness, followed by trustworthiness, followed by dominance. Within each block, three attention check trials were included, where participants were asked to provide a specific rating based on the instructions given by an artificial voice recording (see Materials). There were therefore 369 trials in total ((100 voices + 20 repeated voices + 3 attention checks) * 3 social traits).

Results & Discussion

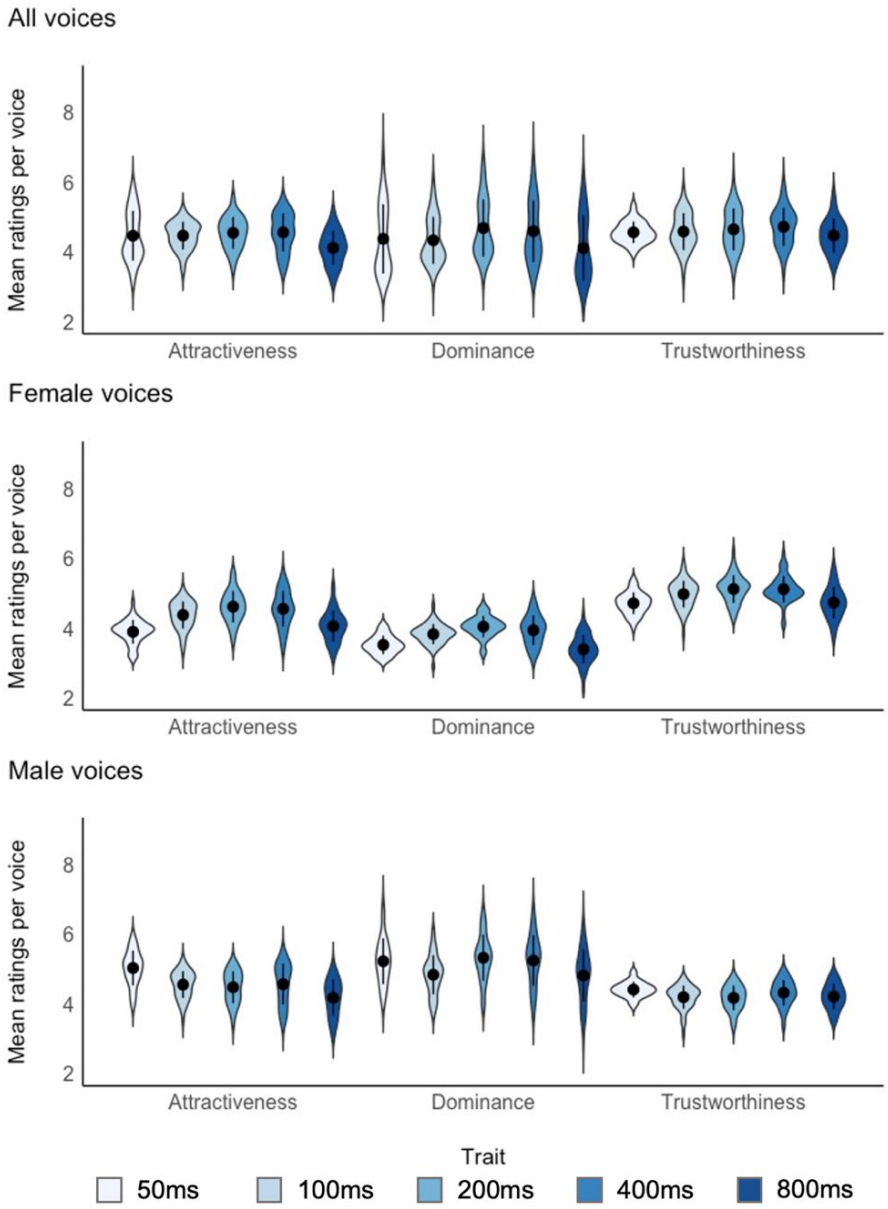
We analyse our data below from a number of perspectives to build a comprehensive picture of how quickly trait impressions can be perceived over time and how these perceived impressions stabilise. We first report measures of inter-rater agreement to establish how well different listeners' ratings are aligned with one another for the different traits and exposure durations. We also analyse the intra-rater agreement to examine whether listeners agree with their own ratings for the voices that were presented twice (see Procedure). We then move on to a different type of analysis, where mean ratings per voice are correlated across exposure durations. Finding significant agreement and/or significant correlations would indicate that responses are not random and that listeners had sufficient exposure to at least start to have perceived a trait impression from the specific exposure duration in question. Finally, we conduct hierarchical regression models for each trait to formally quantify whether and how trait impressions evolve with increasing exposure. For an analysis linking acoustic properties (F0 and formants) to trait impressions, please see Supplementary Analysis 1.

For context, mean ratings by trait and exposure time are visualised across all voices and split by gender in Figure 1 below. From these plots, clear gender effects are apparent, where female voices are rated as overall more trustworthy and less dominant than male voices. No clear gender effects were observed for attractiveness. From Figure 1, it is also apparent that exposure duration appears to somewhat affect the overall trait impressions, although effects differ depending on the gender of the voice: There is some evidence for an early positivity bias for male voices, that is male voices are perceived as numerically more attractive, trustworthy

and dominant (Willis & Todorov, 2006). For female voices, a different pattern emerges, where the shortest exposure duration is, however, rated overall lower, with higher ratings being observed from 100ms, followed by a gradual decrease in mean ratings with increasing exposure.

Figure 1

Mean Ratings per Voice Plotted by Trait and Exposure Duration. Top Panel: Ratings Including All Voices. Middle Panel: Mean Ratings for Female Voices. Bottom Panel: Mean Ratings for Male Voices. Black Dots Show the Mean Rating, Black Lines Show the Standard Deviation.



Inter-Rater Agreement

Two measures of inter-rater agreement were calculated – Cronbach's alpha, being the most commonly used metric in the trait impressions literature, as well as Intraclass Correlation Coefficients (ICCs). This latter measure was used to address some of the criticisms regarding the use of Cronbach's alpha such as relying on assumptions rarely met in psychological research and overestimation of agreement with increasing numbers of raters (Cortina, 1993;

Dunn et al., 2014; Kramer et al., 2018). For the ICC analysis, we used a Two-Way Random model and report the values for absolute agreement, together with 95% confidence intervals. These measures of agreement were calculated separately for each trait and each exposure duration. Data were additionally analysed separately by gender to minimise the possibility of gender differences (see Figure 1) inflating inter-rater agreement. Analyses across genders are reported in Supplementary Analysis 2. A Cronbach's alpha of .70 is generally taken as good agreement, although .60 has also been regarded as acceptable (Taber, 2018). In the current study, we therefore regard acceptable agreement (Cronbach's alpha > .60) as an indication that listeners had sufficient exposure for trait ratings to be well-formed.

When considering Cronbach's alpha for our data, different patterns of results emerge for each trait and voice gender (see Figure 2, as well as Supplementary Tables 3 and 4 for all values of Cronbach's alpha and the ICCs). Broadly speaking, we observe a wide range of agreement in our data, with Cronbach's alpha ranging from -.10 (no agreement) to .88 (very high agreement). There is also some evidence that agreement is generally somewhat lower for shorter exposures. This finding might be evidence that trait impressions from voices do indeed evolve over time (i.e., from 50ms to 400ms of exposure in this study) before stabilising. Notably, alongside other gender effects observed in the data, this pattern of results is most pronounced for female voices, where agreement increases gradually among some fluctuations across the exposure durations (see Figure 2). For male voices, inter-rater agreement tends to increase less between 50ms and 800ms of exposure, which is mostly driven by agreement for shorter exposure durations being higher compared to female voices. As a result, agreement is occasionally relatively stable for male voices across most exposure durations (e.g., for dominance impressions).

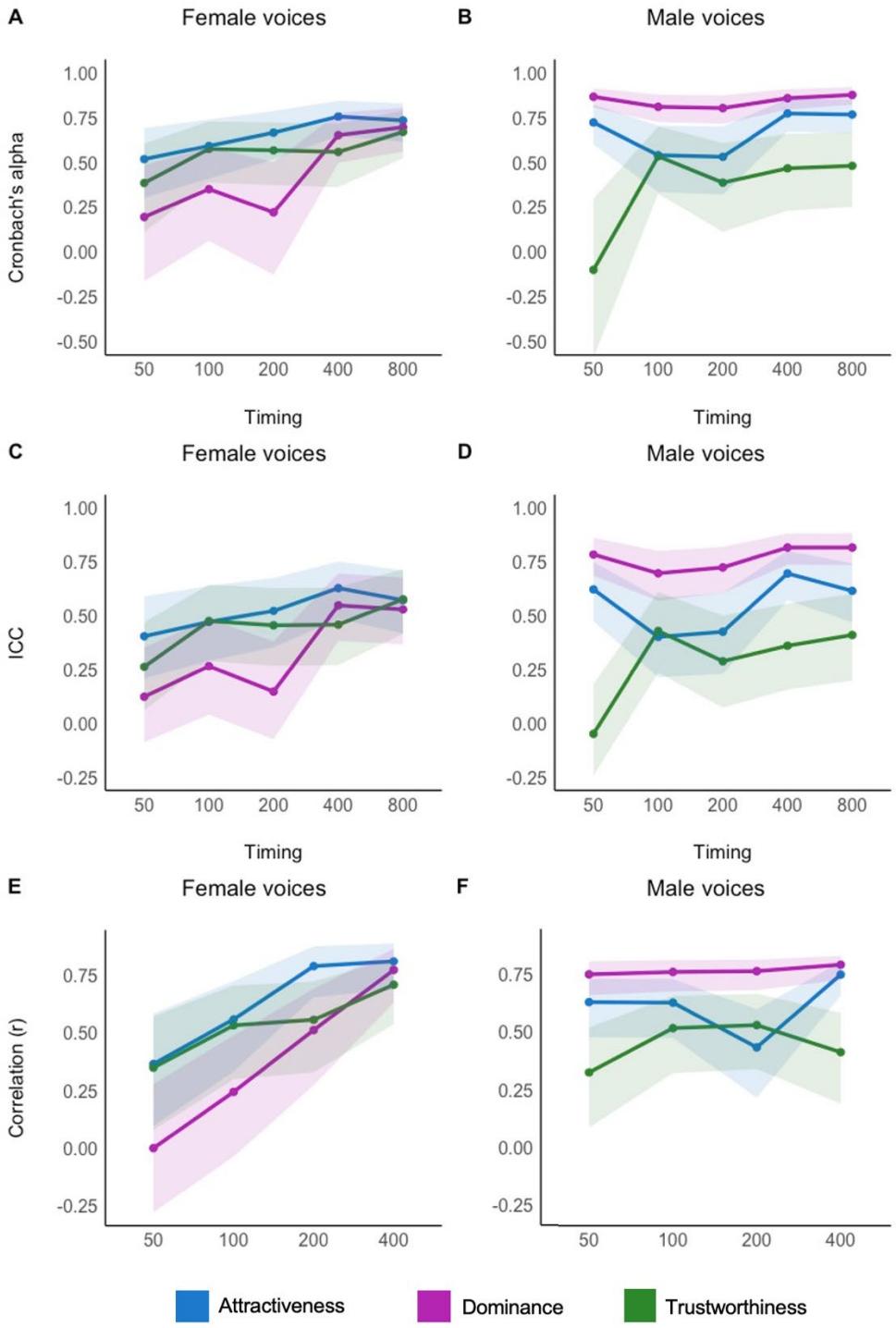
When interpreting our data in light of inter-rater agreement, attractiveness impressions seem to reach acceptable levels (Cronbach's alpha > .60) after 200ms for female voices and after 400ms for male voices. Somewhat surprisingly, attractiveness impressions based on the

shortest exposure duration condition (50ms) were also characterised by good agreement for male voices (Cronbach's alpha = .72) before dropping below our threshold for acceptable agreement for 100ms and 200ms of exposure. For dominance impressions, consistently high agreement across all exposure durations was observed for male voices, while agreement for female voices reached acceptable levels only after 400ms. The overall lowest agreement estimates were seen for impressions of trustworthiness: Impressions of female voices only reached acceptable agreement after 800ms and the highest agreement level for male voices is at .53, thus only approaching the acceptable levels (see Figures 2A and 2B).

Cronbach's alpha, however, has often been criticised as a measure of inter-rater agreement under similar circumstances (Cortina, 1993; Kramer et al., 2018). We therefore additionally provide ICCs as an alternative measure. Overall patterns of results are very similar across ICCs and Cronbach's alpha (see Figures 2 A-D), although ICC values are overall somewhat lower.

Figure 2

Line Graphs Showing Cronbach's Alpha (Panels A+B), ICCs (Panels C+D) And The Correlation Coefficient r (Panels E+F) Across The Different Exposure Durations for Female and Male Voices Separately. Bands Show The 95% CIs Around The Relevant Measure of Interrater Agreement. Colour Figure Is Available In The Online Version Of This Article



Agreement in our study appears to be overall somewhat lower compared to previous studies of trait perception from voices (e.g., Lavan, Mileva & McGettigan, 2021; Lavan, Mileva, Burton, Young, & McGettigan, 2021; Mileva et al., 2020; McAleer et al., 2014). This lower agreement is to be expected since most other studies have used voice stimuli with linguistic content (e.g., words or sentences; but see Rezlescu et al., 2015 who used vowels), while usually also

sampling longer stimulus durations (~390ms in McAleer et al., 2014; stimuli duration of several seconds for Lavan et al., 2021, Lavan, Mileva et al., 2021, and Mileva et al., 2020).

Overall, our analyses, however, show that significant and/or acceptable agreement is reached for most trait impressions after less than 400ms of exposure for female and male voices alike. This significant agreement thus provides initial evidence that even after some very short exposure, perceivers form trait impressions that are not random but do at least to some degree resemble trait impressions of other perceivers. Where no significant agreement was found, trait impressions are either idiosyncratic to each perceiver and/or reflect largely random responses, suggesting that the exposure was too short to form a trait impression. Increasing agreement across exposure durations furthermore suggest that trait impressions may evolve to some degree with increasing exposure.

Intra-rater reliability (item retest)

To tease apart whether the trait impressions that showed non-significant inter-rater agreement reflect random ratings or are the result of valid albeit idiosyncratic trait impressions, we examined the intra-rater reliability (i.e., test-retest reliability) in our data. This measure aims to determine how much perceivers agree with themselves. Despite being an important and meaningful source of variability, intra-rater reliability has been generally neglected in the trait impressions literature. Intra-rater reliability was measured by correlating the first and second ratings of the 20 repeated stimuli (see Procedure), separately for each participant, trait, and exposure duration. Analyses of intra-rater agreement were conducted across gender due to the limited sample size. Average correlations were calculated by first using a Fisher's *r* to *z* transformation, averaging the resulting *z* scores and transforming them back to a correlation using a *z* to *r* transformation.

Table 1

Mean Correlations Between Each Participant's First and Second Rating of the Same Stimuli by Trait and Exposure Duration

Exposure Duration	Attractiveness	Dominance	Trustworthiness
50ms (N=28)	**.49****	**.55****	**.42***
100ms (N=30)	**.39***	**.47****	.34
200ms (N=27)	.37	**.45***	.34
400ms (N=34)	**.51****	**.55*****	**.40***
800ms (N = 32)	**.54****	**.63*****	**.48****

Note. * $p < .05$, ** $p < .01$, *** $p < .001$ significant correlations in **bold**

Table 1 shows the average stimulus test-retest correlations across the three social traits (trustworthiness, dominance, and attractiveness) and the five exposure duration conditions (50ms, 100ms, 200ms, 400ms, and 800ms). Positive correlations are seen for impressions of dominance, demonstrating a degree of intra-rater reliability even with only a 50ms exposure to the voice recordings. Impressions of attractiveness and trustworthiness, however, present a somewhat different pattern of results. Here, we see significant positive correlations in the 50ms condition which then decrease and become significant again after 400ms. These positive correlations thus show that while there is noise in participants' trait impressions, especially after the shorter exposure durations, participants' responses are not fully random. We also note that the non-linear patterns across exposure durations for attractiveness and trustworthiness, however, are only numerical and are not statistically significant ($z = 0.52$, $p = .303$ when comparing the correlations between 50ms and 200ms for attractiveness, $z = 0.34$,

p = .368 when comparing the correlations between 50ms and 100ms for trustworthiness, and z = 0.33, p = .372 when comparing the correlations between 50ms and 200ms for trustworthiness, Eid, Gollwitzer, & Schmidt, 2011). Although statistical significance may not be the most meaningful measure of inference in this case, we nonetheless interpret and treat the non-linear fluctuations across adjacent exposure durations as a reflection of some degree of noise in the data.

Correlation analyses

We next correlated mean ratings per voice of the longest exposure duration (800ms) with mean ratings per voice for the remaining exposure durations, following the analyses of Willis & Todorov (2006) and Todorov et al. (2009). We therefore modelled 800ms as the benchmark at which trait impressions are likely already well-formed (see the time-unconstrained exposure in Willis & Todorov, 2006 and Todorov et al., 2009). As for the inter-rater agreement, we conducted these analyses separately for each trait and voice gender (see Supplementary Figure 2 and Supplementary Table 5 for detailed scatterplots and exact correlation coefficients as well as Supplementary Analysis 3 for correlations across female and male voices). The significance level was corrected for 4 comparisons (alpha = .013). Broadly speaking, moderate-to-strong positive relationships emerged across most of the correlations (see Figures 2E and F for details).

As for the inter-rater agreement, a wide range of correlation strengths were apparent across the different traits and exposure durations. Generally speaking, correlations between 400ms of exposure and 800ms of exposure were strong, ranging between r(48) = .71 - .90 (but see trustworthiness impressions for male voices, r(48) = .47). Furthermore, even for correlations of 50ms and 800ms, moderate positive relationships emerge in most cases.

For attractiveness impressions of female voices, correlations increased in strength from $r(48) = .36$ ($p = .009$) for 50ms to $r(48) = .81$ ($p < .001$) for 400ms. For male voices, strong correlations were apparent across all exposure times, ranging from rs(48) = .50 - .85 (ps < .001), with no clear pattern of increasing correlation strength with increasing exposure duration. There was a substantial decrease in the correlation coefficient for 200ms compared to that for 100ms which was, however, not statistically significant ($z = 1.638$, $p = .051$). This decrease could be driven by the lower intra-rater reliability in this specific condition (see Figure 2). These significant positive correlations suggest that attractiveness impressions based on 50ms of exposure already partially resemble the impressions formed based on 800ms voice samples for both female and male voices. However, for female voices, trait impressions change such that, with increasing exposure, impressions become more similar to the likely well-formed and stable impressions captured by ratings provided for 800ms-long voice samples. Here, the correlations seem to reach some stable level around 200ms.

For dominance impressions of female voices, correlations for 50ms and 100ms were not significant ($r(48) < .01$, $p = .997$ and $r(48) = .24$, $p = .088$). Significant correlations did, however, arise for 200ms ($r(48) = .51$, $p < .001$) and 400ms ($r(48) = .77$, $p < .001$). For male voices, a substantially different picture emerges, with correlations of rs > .85 (ps < .001) being present across all exposure durations. These correlations therefore are similar to our findings for attractiveness impressions, where impressions for female voices emerge gradually with increasing exposure, while this evolution in trait impressions for male voices is much more subtle. Intriguingly, for females, dominance impressions from voice samples of 50ms are entirely unrelated to trait impressions from voice samples of 800ms. This, taken together with the low, non-significant inter-rater agreement at 50ms, suggests both a lack of a shared basis for these dominance impressions as well as a lack of a resemblance to the stable impressions formed at 800ms.

For trustworthiness impressions of female voices, we again observe significant positive correlations of increasing strength across the different exposure durations (see Figure 2E). A similar, albeit less clear trend emerged for male voices, where correlations increase from 50ms to 100ms and 200ms exposure duration ($r(48) = .38$, $p = .013$ vs $r(48) = .59$, $p < .001$ and $r(48) = .60$, $p < .001$). There is a drop in correlation strength for 400ms ($r(48) = .47$, $p < .001$), although this drop is not significant compared to the correlation for 200ms ($z = 0.89$, $p = .187$).

In line with our analyses of inter-rater agreement above, the correlation analyses suggest that 1) from as little as a 50ms exposure to a voice, trait impressions can emerge and that the impressions do indeed resemble well-formed trait impressions (from 800ms of exposure) with only some exceptions. 2) These trait impressions, however, change and evolve over time. 3) Specific patterns depend on the trait rated (see Bar et al., 2006) as well as the gender of the voice. The latter is perhaps best illustrated when looking at the data for dominance ratings, where impressions for female voices only stabilise gradually, while impressions for male voices appear to be well-formed after only 50ms of exposure.

Hierarchical Regressions

To formally quantify whether trait impressions evolve over time to become increasingly similar to the fully-formed impressions observed following 800ms of exposure after each exposure step, we conducted a set of hierarchical multiple regression analyses. Specifically, for each trait and each voice gender, we ran three sets of regression models, where we predicted the mean ratings per voice after 800ms exposure from the mean ratings per voice from one of the shorter exposure duration in a first regression model and compared this model (via R^2 change) to another model that included the same predictor plus the mean ratings from the next longer exposure duration. Thus, for example, mean ratings per voice after 50ms of exposure as the only predictor in Model 1A and the mean ratings per voice after 50ms and 100ms of exposure as predictors in Model 1B (see Table 2). A significant change in R^2 in these models would

suggest that trait impressions do, to some degree, evolve over time; no significant R^2 change would indicate that impressions of voices are stable across the two exposure durations. Since the correlation analysis above has shown that some of the mean ratings are highly correlated, we checked the final models, including all 4 exposure durations for multicollinearity by inspecting the variable inflation scores (VIFs). VIFs were < 10 across all models, with most VIFs not exceeding 2. The highest VIFs were present in the regression models for dominance impressions for male voices, where VIFs ranged from 4.4 to 7.1 – which can still be acceptable (see Field, 2013; Myers, 1990). See Supplementary Analysis 4 for hierarchical regression analyses across female and male voices.

Table 2

Overview of the Hierarchical Regression Analysis, Prediction Item-Wise Mean Ratings After an 800ms Exposure for Each Trait and Voice Gender

Trait	Voice Gender	Predictors	R^2 (A)	R^2 (B)	R^2 Change	*p*
Attractiveness	F	A: 50ms, B: A +100ms	0.13	0.34	0.21	<.001
		A: 100ms, B: A + 200ms	0.31	0.63	0.32	<.001
		A: 200ms, B: A + 400ms	0.62	0.70	0.08	.001
	M	A: 50ms, B: A + 100ms	0.51	0.64	0.13	<.001
		A: 100ms, B: A + 200ms	0.51	0.53	0.02	.212
		A: 200ms, B: A + 400ms	0.25	0.72	0.48	< .001
Dominance	F	A: 50ms, B: A +100ms	<.01	0.07	0.07	.063
		A: 100ms, B: A + 200ms	0.06	0.26	0.20	<.001
		A: 200ms, B: A + 400ms	0.26	0.61	0.35	<.001
	M	A: 50ms, B: A + 100ms	0.72	0.78	0.05	.002
		A: 100ms, B: A + 200ms	0.74	0.79	0.05	.002

		A: 200ms, B: A + 400ms	0.75	0.83	0.08	<.001
Trustworthiness	F	A: 50ms, B: A +100ms	0.12	0.31	0.19	<.001
		A: 100ms, B: A + 200ms	0.28	0.37	0.09	.013
		A: 200ms, B: A + 400ms	0.37	0.53	0.16	<.001
	M	A: 50ms, B: A + 100ms	0.14	0.37	0.23	<.001
		A: 100ms, B: A + 200ms	0.35	0.47	0.12	<.001
		A: 200ms, B: A + 400ms	0.37	0.50	0.13	<.001

In line with the correlation analysis above, we find that even trait impressions formed based on 50ms exposure can significantly predict trait impressions formed after an 800ms exposure (exception: Dominance ratings for female voices). For the remaining steps of the hierarchical regressions, we then observe that with each level of increasing exposure, an additional and significant proportion of the variance in trait impressions based on 800ms exposure can be explained for female and male voices alike with only very few exceptions (see Table 2). This analysis therefore suggests that while trait impressions are established rapidly (as indicated by the agreement measures), they indeed evolve over time with additional exposure adding new information that then shapes the trait impressions. For how long trait impressions are changeable over time, however, appears to depend on the specific traits and on the gender of the perceived voice.

General Discussion

A first impression about a person can be formed both based on the appearance of their face or the sound of their voice. Face and voice impressions share many fundamental properties, such as high inter-rater agreement, an underlying low-dimensional structure, and evidence for real-life consequences (Ballew & Todorov, 2007; McAleer et al., 2014; Mileva et al., 2020;

Oosterhof & Todorov, 2008; Tigue et al., 2012). Another core property of trait impressions from faces and voices that has been highlighted in the literature is how rapidly these trait impressions can be formed. While the rapid time course of trait impressions from faces has been systematically traced, this claim for voices is based on studies finding high agreement using relatively short stimuli (~300-400ms, McAleer et al., 2014; Baus et al., 2019). To our knowledge, there is, however, no work that systematically maps the perception of voice impressions over time. The current study investigated this time course and development of trait impressions from voices via a perceptual gating experiment and assessed how similar the time courses of voice and face trait perception may be.

Trait impressions from voices are formed more slowly than those from faces

Overall, our results show an often-gradual development of voice trait impressions: Inter-rater agreement for trait impressions mostly required up to 400ms of exposure to reach acceptable levels (exception: dominance for male voices). At the same time, there were significant increases in the variance explained in the mean ratings after 800ms of exposure by mean ratings for each consecutive exposure duration (50ms, 100ms, 200ms, and 400ms). Beyond inter-rater agreement, these increases in the variance explained further underline a gradual evolution of trait impressions from voices across time. This is in contrast to what has been reported for face trait impressions, where exposure beyond 100ms does not seem to significantly increase the amount of variability explained in the stable impressions formed after a much longer time period or with no specific time constraints (Willis & Todorov, 2006).

One potential explanation for such differences in the development of trait impressions based on voices vs faces might be the nature of the information carried in the two modalities. While faces provide us with information about both stable structure (such as the presence of a nose in a face) and more transient cues (such as facial expressions), voices mostly include transient information encoded over time. The impact of this potentially more transient and perhaps variable nature of voices on trait perception may be reflected in a recent finding showing that

trait ratings attributed to multiple voice recordings of the same person are much more variable than ratings based on different facial images of the same person (Lavan, Mileva et al., 2021). It is therefore possible that the more transient nature of cues from voices, reflected in the larger within-person variability in trait impressions, leads to listeners' needing additional time to accumulate sufficient information to arrive at a stable trait impression from voices compared to faces.

While information certainly unfolds over time for voices encountered in the real world, it needs to be noted that the current study included sustained vowels, that were in principle (relatively) steady-state. Nonetheless, these voice recordings were produced by humans, such that they are in practice not perfectly static. As such, small fluctuations in the acoustic signal that occur after a certain amount of exposure (e.g., a mild tremor in the voice, a drop in intensity) may have still been present in the voice recordings. Trait impressions could have been to some degree affected by such fluctuations being perceived by listeners: For example, if a decrease in intensity in the voice only occurs after 150ms of exposure, a recording with such a decrease in intensity could receive overall lower dominance ratings after 200ms of exposure compared to 50ms and 100ms of exposure. Similarly, if listeners may not agree on whether the same voice sounds dominant or not after 50ms or 100ms, perceiving such a decrease in intensity may trigger listeners to agree on their dominance impressions. However, for such fluctuations to affect the findings reported in this paper, beyond introducing some noise, fluctuations would need to occur across many of the voice recordings in a systematic and potentially temporally-aligned manner. As a result, the relatively slower time course of trait impressions from voices observed in our study cannot readily be explained by listeners sampling information that substantially evolves over time from the "static" voices. There is, however, still the possibility that trait perception from voices is slower than for faces because participants *expect* that the voice will reveal more information as time passes, while they are perhaps more used to making snap decisions from (static) faces.

Aside from considerations of how the more transient nature of information encoded in voices may shape the time course of trait perception from voices (relative to faces), differences in how different types of socially-relevant judgements (e.g., sex, identity, emotion) are processed from voices vs faces may affect the time courses of trait perception. Some of these socially-relevant judgements appear to be made somewhat independently from faces – for example, sex judgements are usually made faster than identity ones and identity familiarity has been shown to have no impact on the speed of sex judgements (Bruce, 1986). There is, however, evidence that similar judgements are not processed independently from voices, with findings of familiarity enhancing the speed of sex judgements (Burton & Bonner, 2004) and the processing of speech content (Nygaard & Pisoni, 1998). Additionally, voices also encode complex linguistic information that, in most cases, cannot be readily accessed from faces (e.g., most people are poor lip-readers, thus being unable to decode speech from facial information only; Altieri et al., 2011). Given that previous studies have shown that sex, emotion and linguistic information can influence trait perception, we may speculate that another potential explanation for the slower time course of person perception from voices could be found in the differential complexity of how the different types of socially-relevant information interact for voices and faces as well as the differential accessibility of types of information.

As such, our results could have implications for theories focussing on the parallels between general face and voice processing and those suggesting that the voice could be regarded as an “auditory face” (e.g., Belin et al., 2011; Yovel & Belin, 2013). We find that while both face and voice impressions are formed rapidly, their development and evolution follow a distinct pattern for faces vs voices. With voices being inherently dynamic signals, while faces are mostly represented by static images in trait perception studies (although faces tend to be dynamic to some degree outside of experimental studies), our findings highlight a key difference in the development of face and voice impressions. Could it, for example, be the case that face trait impressions are formed quickly and then maintained, while voice trait impressions are more readily updated? This suggests that, while useful under certain

circumstances, a stringent interpretation of the voice being an “auditory face” will be prone to downplay any differences between the two stimulus modalities.

Trait impressions from female voices evolve more gradually than those from male voices

In the following paragraphs, we will further discuss the gender differences we observed for trait perception from voices. Generally, trait impressions were established after less exposure from male than female voices. This pattern of results was particularly evident for ratings of dominance where even 50 milliseconds of exposure were sufficient to form an impression of a male voice, whereas more exposure was needed for female dominance impressions (~400ms). These estimates of the time course were informed not only by the correlations between ratings produced in the shorter exposure duration conditions (50-400ms) and our 800ms exposure condition, but also, critically, following our analysis of intra- and inter-rater agreement, which provided us with important information about the shared nature and the overall reliability of the impressions. While it is possible that a similar pattern of faster trait impressions for male than female identities might emerge for face impressions, this has not been specifically tested in the existing literature with studies generally presenting results across gender (Willis & Todorov, 2006) or collecting impressions of one gender only (Bar et al., 2006). It is, however, notable, that reports of some of the most rapid formation of stable impressions are reported from male faces in particular (e.g., threat and attractiveness, Bar et al., 2006; Rule et al., 2009). We also briefly note that, from the current study, it remains somewhat unclear whether the present pattern of results is partially specific to trait impressions formed by female perceivers, in light of the high proportion of female listeners in our sample and evidence for rater gender differences in trait impressions (Babel, McGuire, & King, 2014; Mattarozzi et al., 2015).

For the current results, we can however, speculate that the different patterns for male and female voices from even the shortest exposure durations show that trait impressions and

gender perception are closely linked. Studies have already shown that face impressions are driven by social categorisation and overgeneralisation processes (Secord, 1958; Zebrowitz & Montepare, 2008). If gender perception influences trait perception for voices, a trait impression could, in principle, start to be formed as soon as a gender judgement has been made. For face perception, South Palomares and Young (2018) already demonstrate a similar, and rapid time course for impression formation and gender classification. With the human voice, and its pitch in particular, being sexually dimorphic (Titze, 2000), gender discrimination from voices can be achieved with 99% accuracy after 50ms exposure to vowel segments (Whiteside, 1998) and with 75% accuracy after less than a 30ms of exposure (Owren, Berkowitz, & Bachorowski, 2007). Intriguingly, studies looking at the time course of gender classification decisions demonstrate that male voices can be classified faster and more accurately than female voices (Lass et al., 1976; Coleman, 1971; Owren et al., 2007). Thus, assuming that gender classification is closely linked to trait perception, it is possible that the slower development and lower agreement we observe for trait impressions from female voices may partially reflect the time course differences in gender classification decisions. Furthermore, Latinus and Taylor (2012) present two ERP studies focussing on the perception of voice gender and the specific role of vocal pitch. Their results suggest that pitch information is picked up from the voice shortly after stimulus onset (at between 30-87ms) whereas “true” gender processing occurs later on at approximately 200ms. As pitch could be used as a rapid index of gender, this could explain both how we can form a stable first impression within 50ms (in some cases) and why some impressions evolve further with additional exposure time.

Going beyond overgeneralisation effects, we may also speculate that the observed gender differences could also provide us with an example of the dual-route processing suggested by Over and Cook (2018). Based on dual-process theories of social and more general human cognition (Frankish, 2010; Kahneman, 2003; Kruglanski & Orehek, 2007), they propose two distinct processing routes underpinning trait perception from (unfamiliar) faces: an intuitive (automatic) route and a more controlled (reflective) route that requires more deliberate

reasoning about an individual's appearance and behaviour (see also Campbell-Kibler, 2016). These two routes likely work independently from one another, however, there is scope and evidence for some level of interaction between them. For example, more reflective top-down processes might override automatic judgements (such as stereotypes, Devine, 1989) and automatic judgments can introduce biases in controlled processes (e.g., an initial negative evaluation might require a larger amount of subsequent positive information in order to change the overall perception of a person, Over & Cook, 2018). Taking dominance impressions again as an example – upon hearing a voice, listeners might be making a fast and automatic gender judgement (likely based on pitch information) which then leads to prioritising the processing of male (low-pitched) voices via the intuitive, fast route given that from an evolutionary perspective, threat is more likely to be encountered from male identities. The processing of female (higher-pitched) voices might not require the same high priority route and dominance impressions from these voices might therefore take longer to stabilise and can allow for more information to be processed in order to arrive at a stable judgement.

Trait impressions are formed most quickly for dominance

We also observe differences in the way impressions of the three fundamental traits are formed: Attractiveness and trustworthiness ratings for male and female voices alike, as well as dominance impressions for female voices require several hundreds of milliseconds to form and stabilise, while dominance impressions for male voices emerge within 50ms of exposure. This pattern of results is closely replicated when looking across male and female voices (see Supplementary Materials), where dominance impressions are also formed rapidly within 50ms of exposure, while attractiveness and trustworthiness ratings evolve more gradually. The rapid emergence of dominance impressions is a potentially intriguing finding because it shows that, compared to the speed of face trait ratings, dominance impressions might be formed faster from the voice, whereas attractiveness and trustworthiness impressions are potentially formed faster (or with broadly similar speed) from the face. These findings then fit nicely with studies of audiovisual trait perception reporting that the information from the voice is more important

than that from the face for audiovisual impressions of dominance (Mileva et al., 2018; Rezlescu et al., 2015). A different pattern is observed for impressions of trustworthiness, where face information either is more important (Mileva et al., 2018) or contributes equally to audiovisual trustworthiness impressions (Rezlescu et al., 2015). Similarly, Rezlescu et al. (2015) also report that face information is more important for judgements of attractiveness. These differences in the relative importance of faces and voices for audiovisual trait impressions therefore appear to follow the speed at which these judgements are formed from the face and the voice.

Conclusion and future directions

Overall, our data thus provide a detailed first examination of the formation and development of trait impressions from voices. However, many questions regarding the time course of vocal trait perception remain unanswered: In the current study, we used sustained vowels as stimuli. These voice samples were specifically chosen to ensure that any effects observed in the current study were driven by the additional voice exposure, rather than by differences in how phonetic and linguistic content unfolds across the different exposure times. As such, trait impressions formed from these stimuli can only be based on the basic vocal characteristics of the voice, with little meaningful linguistic (and paralinguistic) information being available. As noted above, the time course of trait perception from voice recordings that include linguistic information may be different and potentially more prolonged than what we observed in the current study. Given that linguistic information generally unfolds relatively slowly over time compared to e.g., gender perception (the short, two-syllable word “Hello” takes ~390ms to be produced in full, see McAleer et al., 2014), it is likely that using stimuli that include linguistic information per se would not change how quickly trait impressions are initially formed based on the vocal characteristics of a person. Rather, linguistic content, such as the meaning of the words spoken, might affect trait impressions based on voices in line with the message conveyed once this message has been perceived. As such, a person might be initially perceived as trustworthy due to their vocal characteristics before listeners have processed the

linguistic content of a potentially offensive utterance, which should lower the perceived trustworthiness. Similarly, there is also a wealth of sociolinguistic literature on the effects of e.g., accent perception on how listeners evaluate and react to other people in a valenced or prejudiced manner. These studies often also use much longer voice samples (e.g., Bayard et al., 2001), although similar effects have been reported for short samples (~400ms, Purnell et al., 1999 for accent perception).

In addition to the linguistic information that voices usually carry in the real world, our study also did not account for the wealth of paralinguistic information, i.e. information conveyed by *how* things are being said, that is usually encoded in real-world voices: One such type of paralinguistic information is emotional content, which has already been shown to affect trait perception from faces (Hess et al., 2000; Krumhuber et al., 2007; Montepare & Dobish, 2003; Mueser et al., 1984) and for voices (Pinheiro et al., 2021). Pell and Kotz (2011) have furthermore outlined the time course of emotion perception from speech, reporting that different emotions are associated with recognition time courses: While all emotions can be reliably recognised from under 2 seconds of exposure, anger appears to be recognised sooner than happiness, after 710ms vs 977ms respectively. As with linguistic information, paralinguistic information is thus likely to further colour trait impressions from voices. This potentially happens on a somewhat slower timescale than e.g. the perception of voice quality.

Thus, while the current study has for the first time outlined the time course of trait impressions from voices, from their initial inception to reaching relative stability, the formation of trait impressions from more naturalistic stimuli is likely to be much more complex, with different sources of information encoded in the voices interacting and unfolding on different time scales. In light of these many sources of information that are routinely available from voices in naturalistic settings, much further work is needed to explore and model how trait impressions are shaped *after* their initial formation, over the course of the first few seconds of meeting a person. Nevertheless, our data provide important methodological and theoretical insights

about the time course of social evaluation processes as well as the parallels (and some differences) between face and voice perception in general.

References

Altieri, N. A., Pisoni, D. B., & Townsend, J. T. (2011). Some normative data on lip-reading skills (L). *The Journal of the Acoustical Society of America*, *130*(1), 1-4. https://doi.org/10.1121/1.3593376.

Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. PloS One, 9(2), e88616. https://doi.org/10.1371/journal.pone.0088616

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports, 9*(1), 414. https://doi.org/10.1038/s41598-018-36518-6

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. (2001). Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics, 5*(1), 22–49. https://doi.org/10.1111/1467-9481.00136

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception, 15*(4), 387–397. https://doi.org/10.1068/p150387

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., & Bonner, L. (2004). Familiarity influences judgments of sex: The case of voice recognition. *Perception, 33*(6), 747–752. https://doi.org/10.1068/p3458

Campbell-Kibler, K. (2016). Toward a cognitively realistic model of meaningful sociolinguistic variation. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151). Cambridge: Cambridge University Press.

Chen, D. L., Kumar, M., Motwani, V., & Yeres, P. (2016). Is justice really blind? And is it also deaf? *SSRN*. http://dx.doi.org/10.2139/ssrn.2816567

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research, 14*(3), 565–577. https://doi.org/10.1044/jshr.1403.565

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden*. Weinheim, Germany: Beltz.

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62–73. https://doi.org/10.1016/j.tourman.2016.01.013

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass, 5*(10), 914–926. https://doi.org/10.1111/j.1747-9991.2010.00330.x

Hehman, E., Xie, S. Y., Ofosu, E. K., & Nespoli, G. (2018). Assessing the point at which averages are stable: A tool illustrated in the context of person perception. https://psyarxiv.com/2n6jq/

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior, 24*(4), 265–283. https://doi.org/10.1023/A:1006623213355

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Klofstad, C. A. (2016). Candidate voice pitch influences election outcomes. *Political Psychology, 37*(5), 725–738. https://doi.org/10.1111/pops.12280

Klofstad, C. A., & Anderson, R. C. (2018). Voice pitch predicts electability, but does not signal leadership ability. *Evolution and Human Behavior, 39*(3), 349–354. https://doi.org/10.1016/j.evolhumbehav.2018.02.007

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences, 279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Kramer, R. S. S., Mileva, M., & Ritchie, K. L. (2018). Inter-rater agreement in trait judgements from faces. *PloS One, 13*(8), e0202655. https://doi.org/10.1371/journal.pone.0202655

Kruglanski, A. W., & Orehek, E. (2007). Partitioning the domain of social inference: Dual mode and systems models and their alternatives. *Annual Review of Psychology, 58*, 291–316. https://doi.org/10.1146/annurev.psych.58.110405.085629

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*(4), 730–735. https://doi.org/10.1037/1528-3542.7.4.730

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America, 59*(3), 675–678. https://doi.org/10.1121/1.380917

Latinus, M., & Taylor, M. J. (2012). Discriminating male and female voices: differentiating pitch and gender. *Brain Topography, 25*(2), 194–204. https://doi.org/10.1007/s10548-011-0207-9

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review, 26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lavan, N., Mileva, M., Burton, A. M., Young, A. W., & McGettigan, C. (2021). Trait evaluations of faces and voices: Comparing within-and between-person variability. *Journal of Experimental Psychology: General, 150*(9), 1854–1869. https://doi.org/10.1037/xge0001019

Lavan, N., Mileva, M., & McGettigan, C. (2021). How does familiarity with a voice affect trait judgements? British Journal of Psychology, 112(1), 282–300. https://doi.org/10.1111/bjop.12454

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience, 30*(22), 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PloS One, 13*(10), e0204991. https://doi.org/10.1371/journal.pone.0204991

Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of gender and personality on first impression. *PloS One, 10*(9), e0135529. https://doi.org/10.1371/journal.pone.0135529

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PloS One, 9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance, 44*(1), 128–138. https://doi.org/10.1037/xhp0000439

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2020). The role of face and voice cues in predicting the outcome of student representative elections. *Personality and Social Psychology Bulletin, 46*(4), 617–625. https://doi.org/10.1177/0146167219867965

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*(4), 237–254. https://doi.org/10.1023/A:1027332800296

Mueser, K. T., Grau, B. W., Sussman, S., & Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology, 46*(2), 469–478. https://doi.org/10.1037/0022-3514.46.2.469

Myers, R. H. (1990). *Classical and modern regression with applications (Vol. 2)*. Belmont, CA: Duxbury press.

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44*(3), 342–366. https://doi.org/10.1006/brcg.1999.1196

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376. https://doi.org/10.3758/BF03206860

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition, 170*, 190–200. https://doi.org/10.1016/j.cognition.2017.10.002

Owren, M. J., Berkowitz, M., & Bachorowski, J. A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics, 69*(6), 930–941. https://doi.org/10.3758/BF03193930

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One, 6*(11), e27256. https://doi.org/10.1371/journal.pone.0027256

Pinheiro, A. P., Anikin, A., Conde, T., Sarzedas, J., Chen, S., Scott, S. K., & Lima, C. F. (2021). Emotional authenticity modulates affective and social trait inferences from voices. *Philosophical Transactions of the Royal Society B, 376*(1840), 20200402. https://doi.org/10.1098/rstb.2020.0402

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*(1), 10–30. https://doi.org/10.1177/0261927X99018001002

Pützer, M. & Barry, W. J. (n.d.). Saarbrücker Voice Database, Available at http://stimmdb.coli.uni-saarland.de/.

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior, 39*(4), 355–370. https://doi.org/10.1007/s10919-015-0214-8

Rule, N. O., Ambady, N., & Adams Jr, R. B. (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception, 38*(11), 1688–1699. https://doi.org/10.1068/p6384

Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience, 13*(1), 1–13. https://doi.org/10.1093/scan/nsx142

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(1), 15–25. https://doi.org/10.1002/wcs.1261

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), Person Perception and Interpersonal Behavior (pp. 300–315). Stanford, CA: Stanford University Press.

South Palomares, J. K., & Young, A. W. (2018). Facial first impressions of partner preference traits: Trustworthiness, status, and attractiveness. *Social Psychological and Personality Science, 9*(8), 990–1000. https://doi.org/10.1177/1948550617732388

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology, 49*(4), 771–775. https://doi.org/10.1016/j.jesp.2013.02.003

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition, 127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research from Science Education, 48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Titze, I. R. (2000). Principles of voice production (second printing). Iowa City, IA: National Center for Voice and Speech, 229–233.

Todorov, A. (2017). Face value. Princeton University Press.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. https://10.1126/science.1110589

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Whiteside, S. P. (1998). Identification of a speaker's sex: A study of vowels. *Perceptual and Motor Skills, 86*(2), 579–584. https://doi.org/10.2466/pms.1998.86.2.579

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences, 24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*(6), 263–271. https://doi.org/10.1016/j.tics.2013.04.004

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

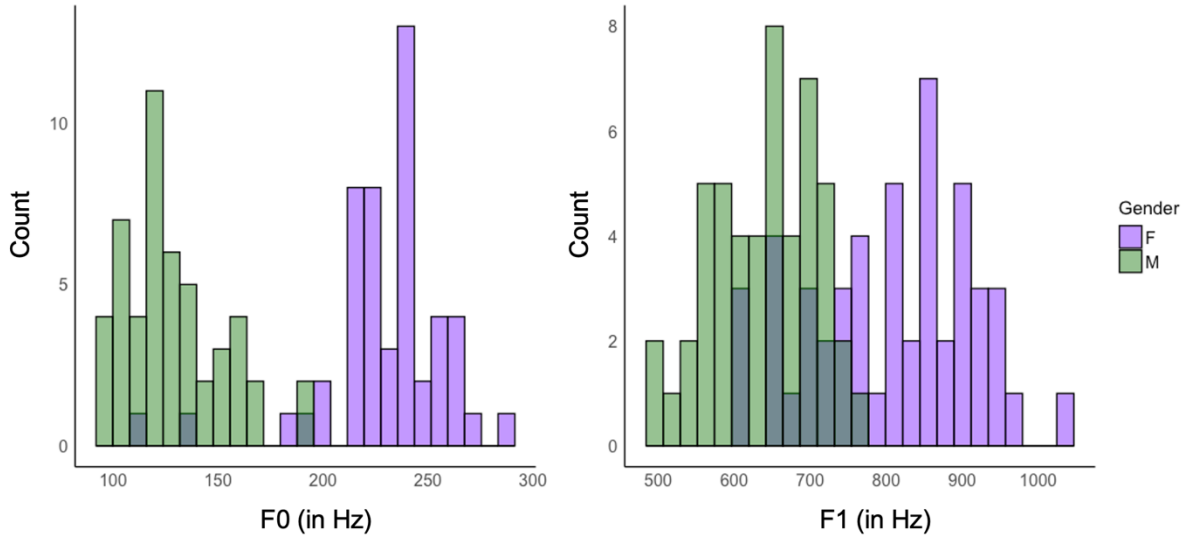
Supplementary Materials

Supplementary Analysis 1: The relationship between mean ratings per item and acoustics

We extracted acoustic features, specifically F0, corresponding to voice pitch and F1, the first formant, extracted here to be a proxy for vocal tract length (VTL). The mean F0 was 229.5Hz (SD: 30.0 Hz) for female voices and 127.7Hz (SD: 24.7 Hz) for male voices. The F1 was 806.5Hz (SD: 105.4 Hz) for female voices and 639.9Hz (SD: 24.7 Hz) for male voices (see Supplementary Figure 1 below for the distribution of F0 and F1 values).

Supplementary Figure 1

Histograms Showing the Distributions of F0 and F1 Values for Male and Female Voices



Supplementary Table 1

Correlations Between Item-Wise Mean Ratings by Trait, Gender, and Exposure Time with Two Acoustic Measures: F0 (Top Table) and F1 (Bottom Table)

F0

Exposure Duration	Attractiveness			Dominance			Trustworthiness		
	All voices	F	M	All voices	F	M	All voices	F	M
50ms	**-.82*****	-0.17	**-.50*****	**-.88*****	**-.45****	**-.48*****	**.53*****	**.47*****	-0.01
100ms	**-.29****	.24	**-.47*****	**-.77*****	-.12	**-.38****	**.71*****	0.17	0.05
200ms	.15	.29*	**-.22**	**-.78*****	-.16	**-.42****	**.79*****	**.37****	0.16
400ms	-.10	.05	**-.46*****	**-.73*****	-.11	**-.38****	**.67*****	0.01	0.16
800ms	-.15	.23	**-.44****	**-.76*****	-.06	**-.48*****	**.5*****	0.05	0.12

F1 (proxy for vocal tract length)

Exposure Duration	Attractiveness			Dominance			Trustworthiness		
	All voices	F	M	All voices	F	M	All voices	F	M
50ms	-.22	-.22	**-.29***	**-.66*****	-.09	-.23	**.47*****	**.42*****	-.22
100ms	**.31***	**.31***	-.07	**-.54*****	.10	-.21	**.69*****	**.47*****	.20
200ms	**.36****	**.26****	.16	**-.55*****	-.02	-.15	**.71*****	**.54*****	.24
400ms	**.29***	**.29***	.01	**-.44*****	.22	.02	**.50*****	.06	-.03
800ms	**.29***	**.29***	-.16	**-.40*****	**.44*****	.07	**.42*****	.19	-.05

Note. F = female voices, M = male voices; * $p < .05$, ** $p < .01$, *** $p < .001$, significant correlations in **bold**, negative correlations in red, positive correlations in green

These kinds of acoustic measures have been previously shown to be influential for trait impressions from voices (McAleer et al., 2014). As can be seen from Supplementary Table 1, we observe consistent and significant negative relationships between F0 and attractiveness and dominance ratings for male voices: The lower the pitch, the more attractive and dominant male voices are perceived to be. No such relationship emerges for trustworthiness, although most correlations are weakly positive. For female voices, no consistent significant relationships between F0 and trait ratings emerge, if anything, there are trends suggesting that female voices that are higher in F0 are perceived as more attractive and trustworthy while female voices that are lower in F0 are perceived as more dominant. When looking across both male and female voices, the effects of gender on trait impressions becomes apparent: We observe consistently strong negative correlations for dominance impressions, confirming that male voices (which are usually lower in F0 than female voices) are perceived as overall more dominant. Conversely, we observe a moderate-to-strong positive correlation between F0 and trustworthiness impressions, reflecting that female voices (which are usually higher in F0 than male voices) are perceived as overall more trustworthy than male voices.

For correlations with F1, as a proxy for vocal tract length, fairly consistent relationships emerge for attractiveness and trustworthiness ratings for female voices: Higher F1 is associated with shorter vocal tract lengths. As such, female voices are perceived as more trustworthy and attractive when the voice overall sounds smaller in terms of the voice quality. No clear relationship between F1 and dominance impressions for female voices was apparent. Furthermore, there were also no consistent relationships between F1 and trait impressions for male voices. Looking

across female and male voices, we can again observe the effects of gender on the correlations between F1 on dominance ratings: Males tend to have longer vocal tracts and thus lower F1s. As a result, we observe a negative relationship between voices F1 and dominance ratings. For trustworthiness (and to a lesser degree for attractiveness) impressions, we observe the opposite pattern, largely reflecting that females are perceived to be more trustworthy.

Supplementary Analysis 2: Inter-rater agreement for female and male voices combined

Trait impressions of attractiveness based on both male and female voices converge on acceptable agreement at 400ms, such that no systematic increase in agreement is observed after this exposure time (see Supplementary Table 2). However, surprisingly, trait impressions based on the shortest exposure time (50ms) also already exhibit high agreement. Dominance impressions across both male and female voices show very high inter-rater agreement even at the shortest exposure time. For trustworthiness impressions, acceptable agreement is reached after 100ms, with agreement being overall relatively stable across all exposure durations other than 50ms, where agreement is quite low.

Supplementary Table 2

Inter-Rater Agreement (Cronbach's Alpha and Intraclass Correlation Coefficients, with 95% Confidence Intervals) for Male and Female Voices Combined.

Trait	Exposure Duration	Cronbach's Alpha [95% CIs]	ICC [95% CIs]
Attractiveness	50ms	.83 [.78, .87]	**.81***** [.75, .86]
	100ms	.47 [.30, .61]	**.40***** [.24, .54]
	200ms	.47 [.30, .61]	**.42***** [.26, .56]
	400ms	.70 [.61, .78]	**.64***** [.54, .73]
	800ms	.66 [.56, .75]	**.57***** [.45, .67]
Dominance	50ms	.93 [.90, .95]	**.90***** [.87, .93]
	100ms	.84 [.79, .88]	**.79***** [.73, .85]
	200ms	.86 [.82, .90]	**.82***** [.76, .87]
	400ms	.90 [.86, .92]	**.88***** [.84, .91]
	800ms	.91 [.89, .94]	**.88***** [.84, .91]
Trustworthiness	50ms	.20 [-.04, .41]	**.16*** [-.03, .34]
	100ms	.76 [.68, .82]	**.71***** [.62, .78]
	200ms	.78 [.71, .84]	**.71***** [.63, .79]
	400ms	.74 [.67, .81]	**.68***** [.59, .76]
	800ms	.65 [.54, .74]	**.61***** [.50, .71]

Note. * $p < .05$, *** $p < .001$, significant ICCs in **bold**

Supplementary Table 3

Inter-Rater Agreement (Cronbach's Alpha and Intraclass Correlation Coefficients, with 95% Confidence Intervals) for Female Voices.

Trait	Exposure Duration	Cronbach's Alpha [95% CIs]	ICC [95% CIs]
Attractiveness	50ms	.52 [.30, .69]	**.40***** [.06, .47]
	100ms	.59 [.41, .74]	**.47***** [.29, .64]
	200ms	.67 [.52, .79]	**.52***** [.35, .68]
	400ms	.76 [.65, .84]	**.63***** [.49, .75]
	800ms	.73 [.62, .83]	**.57***** [.42, .71]
Dominance	50ms	.20 [-.16, .49]	.12 [-.09, .35]
	100ms	.35 [.06, .58]	**.27**** [.04, .49]
	200ms	.22 [-.13, .50]	.15 [-.08, .38]
	400ms	.65 [.50, .78]	**.55***** [.38, .70]
	800ms	.70 [.56, .81]	**53***** [.37, .68]
Trustworthiness	50ms	.39 [.11, .61]	**.26**** [.06, .47]
	100ms	.58 [.39, .73]	**.48***** [.29, .64]
	200ms	.57 [.37, .72]	**.46***** [.27, .63]
	400ms	.56 [.36, .72]	**.46***** [.27, .63]
	800ms	.67 [.52, .79]	**.58***** [.42, .72]

Note. ** $p < .01$, *** $p < .001$, significant ICCs in **bold**

Supplementary Table 4

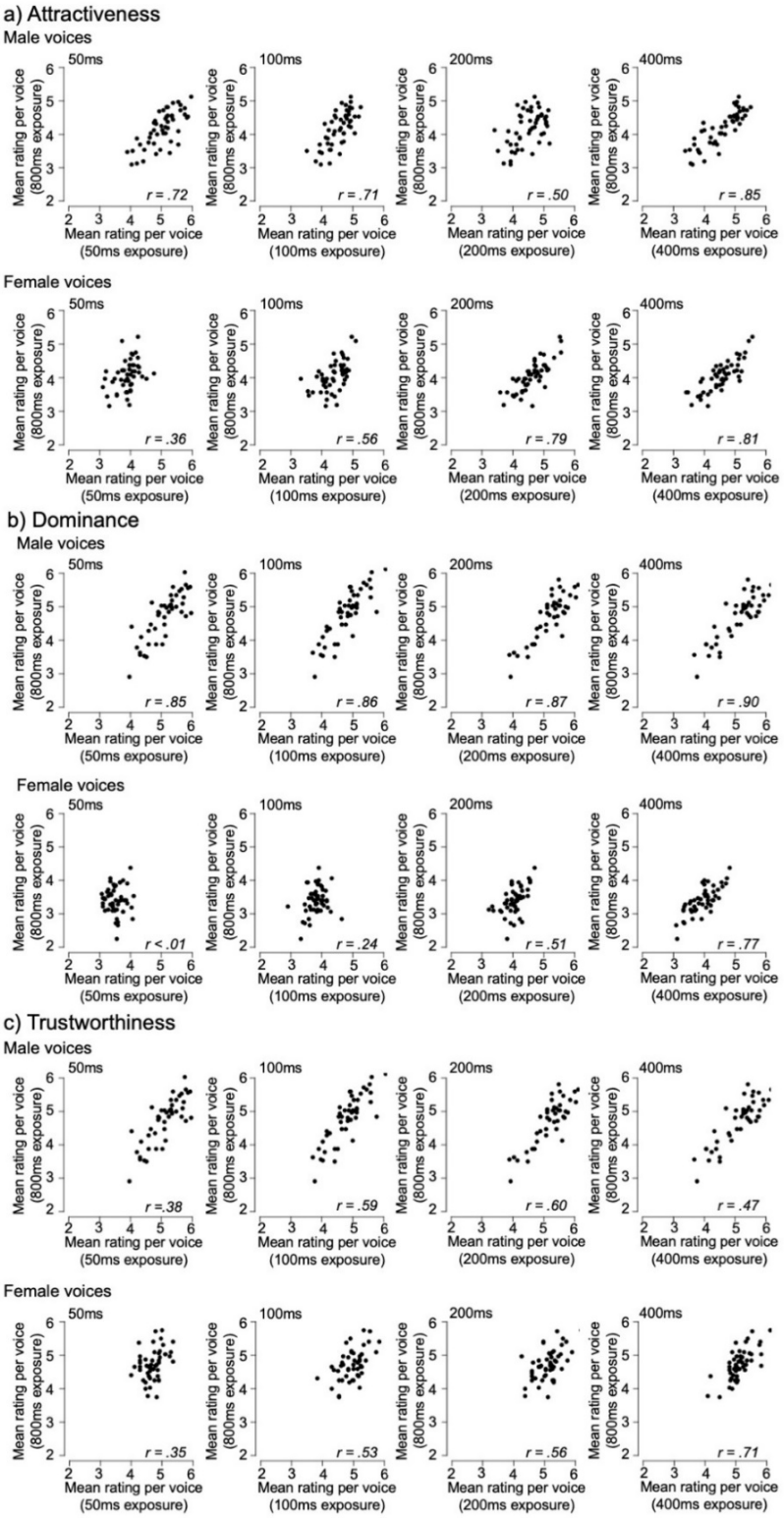
Inter-Rater Agreement (Cronbach's Alpha and Intraclass Correlation Coefficients, with 95% Confidence Intervals) for Male Voices.

Trait	Exposure Duration	Cronbach's Alpha [95% CIs]	ICC [95% CIs]
Attractiveness	50ms	.72 [.60, .82]	**.62***** [.47, .75]
	100ms	.54 [.34, .71]	**.40***** [.21, .58]
	200ms	.53 [.32, .70]	**.43***** [.23, .61]
	400ms	.77 [.67, .85]	**.70***** [.57, .80]
	800ms	.77 [.66, .85]	**.62***** [.47, .74]
Dominance	50ms	.87 [.81, .91]	**.78***** [.69, .86]
	100ms	.81 [.73, .88]	**.70***** [.57, .80]
	200ms	.80 [.72, .87]	**.72***** [.61, .82]
	400ms	.86 [.80, .91]	**.82***** [.74, .88]
	800ms	.88 [.82, .92]	**.82***** [.74, .88]
Trustworthiness	50ms	-.10 [-.59, .30]	-.05 [-.24, .18]
	100ms	.53 [.33, .70]	**.43***** [.24, .61]
	200ms	.39 [.11, .61]	**.29**** [.07, .50]
	400ms	.47 [.23, .66]	**.36***** [.16, .56]
	800ms	.48 [.25, .67]	**.41***** [.20, .60]

Note. ** $p < .01$, *** $p < .001$, significant ICCs in **bold**

Supplementary Figure 2

Scatterplots of the Relationships Between Item-Wise Mean Ratings for Trait Impressions After 800ms and the Shorter Exposure Durations. Data are Plotted by Trait and Voice Gender.



Supplementary Table 5

Correlation Coefficients for Correlations Between Item-Wise Mean Ratings After an 800ms Exposure and Those for the Remaining Shorter Exposure Durations by Trait and Voice Gender

Trait	Exposure Duration	Female	Male
Attractiveness	50ms	**.36**	**.72**
	100ms	**.56**	**.71**
	200ms	**.79**	**.50**
	400ms	**.81**	**.85**
Dominance	50ms	<.01	**.85**
	100ms	.24	**.86**
	200ms	**.51**	**.87**
	400ms	**.77**	**.90**
Trustworthiness	50ms	.35	**.38**
	100ms	**.53**	**.59**
	200ms	**.56**	**.60**
	400ms	**.71**	**.47**

Note. Significant correlations after multiple comparison correction in **bold**

Supplementary Analysis 3: Correlations between 800ms and shorter durations for female and male voices combined

Supplementary Table 6 shows the correlations between item-wise mean ratings for different exposure durations (50ms, 100ms, 200ms, 400ms) and 800ms when looking across female and male voices. We find strong positive correlations across all

exposure durations and traits, with these effects most likely being driven by underlying gender differences in trait impressions for male vs female voices. These data clearly show a distinct pattern of development for each trait. Dominance judgements seem to be well formed, even at 50ms, whereas trustworthiness judgements require a little longer (100ms). The development of attractiveness judgements seems to be much more gradual and only approaches stability at 400ms.

Supplementary Table 6

Correlation Coefficients for Correlations Between Item-Wise Mean Ratings after an 800ms Exposure and the Remaining Shorter Exposure Durations by Trait for Male and Female Voices Combined

Trait	Exposure Duration	Correlation Coefficient
Attractiveness	50ms	.45***
	100ms	.65***
	200ms	.60***
	400ms	.83***
Dominance	50ms	.89***
	100ms	.88***
	200ms	.92***
	400ms	.94***
Trustworthiness	50ms	.52***
	100ms	.71***
	200ms	.72***
	400ms	.74***

Note. *** $p < .001$

Supplementary Analysis 4: Hierarchical regression analysis for female and male voices combined

Supplementary Table 7

Overview of the Hierarchical Regression Analysis, Prediction Item-Wise Mean Ratings after 800ms Exposure for Each Trait for Female and Male Voices Combined

Trait	Predictors	R^2 (A)	R^2 (B)	R^2 Change	*p*
Attractiveness	A: 50ms, B: A + 100ms	0.20	0.44	0.24	<.001
	A: 100ms, B: A + 200ms	0.42	0.51	0.09	<.001
	A: 200ms, B: A + 400ms	0.36	0.69	0.33	<.001
Dominance	A: 50ms, B: A + 100ms	0.78	0.82	0.03	<.001
	A: 100ms, B: A + 200ms	0.78	0.85	0.07	<.001
	A: 200ms, B: A + 400ms	0.84	0.91	0.07	<.001
Trustworthiness	A: 50ms, B: A + 100ms	0.27	0.53	0.25	<.001
	A: 100ms, B: A + 200ms	0.51	0.57	0.06	<.001
	A: 200ms, B: A + 400ms	0.52	0.61	0.09	<.001

For completeness, we present the hierarchical regression analysis reported for female and male voices separately in the main text here when considering all voices together in the same analysis (see Supplementary Table 7). The picture that emerged from these three hierarchical regression analyses (one per trait) is similar to what we observed when splitting the data by gender: With increasing exposure duration,

modelled here via additional predictors, unique and significant portions of variance are explained with each increase in exposure duration. Final models, account for most of the variance in trait impressions based on 800ms exposure (61% - 91%) for all traits.