How trustworthy is your voice? The effects of voice manipulation on the perceived trustworthiness of novel speakers

Bibi Boehme

Supervised by: Prof Pascal Belin

March 2014

How trustworthy is your voice? The effects of voice manipulation on the perceived trustworthiness of novel speakers

ABSTRACT

A person's voice is not only loaded with cues to age, sex and emotional state, listeners also readily form personality impressions of novel speakers. Based on research on face perception, suggesting that people rapidly and reliably evaluate faces on personality traits which can be summarized in a two-dimensional space, with one dimension emphasizing warmth/likability/trust and the other emphasizing strength/dominance, a similar model has been proposed for personality impressions from voices. The present study builds upon these findings, investigating how trustworthiness is perceived in voices and conveyed by novel speakers. For both male and female voices, morph continua were created between voices previously rated low/high on trustworthiness to examine whether the manipulation towards an averaged un-/trustworthy voice would shift perception of the voices towards un-/trustworthiness, respectively, and whether vocal caricatures of these prototypical voices would enhance the effect. Through an online rating experiment, 422 participants rated 18 voices on their trustworthiness. Akin to a 'zero acquaintance' scenario, the stimuli were sub-second vocal utterances of a single word and no contextual information was provided. Repeated measures ANOVAs and correlation analyses showed that, for male voices, there was a positive linear relationship between the responses and manipulations towards a trustworthiness caricature. For female voices, on the other hand, the results were inconclusive due to an error in the generation of the stimuli. The findings contribute to the empirical bases for understanding personality judgments from brief speech signals and are discussed in light of previous research, implications, potential applications, limitations and suggestions for further research.

Key words:	Social perception	Trustworthiness	Voice	Personality trait	First impressions

Introduction

Personality impressions from voices

A person's voice is not only loaded with cues to age, gender and affective state (Belin et al., 2011; Yovel & Belin, 2013), it has also intrigued researchers to investigate how personality can be inferred from voice. Early research in this area was inspired by the development of radio broadcasting (e.g. Allport & Cantril, 1934), but associated with methodological problems, including a lack of attention to speaker differences and the exclusive use of monologues (Kramer, 1964). In an attempt to overcome these, Scherer (1972) emphasized the importance of factors such as the existence of stable voice personality relationships and the listener's ability to isolate and perceive accurately the relevant vocal cues. Yet, he failed to specify how these aspects play a role. More recent studies have focused on specific personality traits, including attractiveness and dominance (Berry, 1990; Bruckert et al., 2010; Feinberg et al., 2012; Hughes et al., 2004; Zebrowitz-McArthur & Montepare, 1989; Zuckerman & Driver, 1989).

Similarities between face and voice processing

Although the study of personality judgements from voices has been of interest for many years, research has primarily focused on monologues and extended speech (Zuckerman & Driver, 1989; Montepare & Zebrowitz-McArthur, 1987; Berry, 1990; Zebrowitz-McArthur & Montepare, 1989).
The current study, though, is inspired by similarities in face and voice processing (Bruce & Young, 1986; Belin et al., 2004, 2011; Yovel & Belin, 2013).
Studies on face perception have shown that people rapidly and reliably evaluate faces on a number of personality traits (Willis & Todorov, 2006; Bar et al., 2006; Todorov et al., 2008; Zebrowitz & Montepare, 2008). Yet, whereas face research on first impressions has focused on brief exposure to static pictures, establishing similar conditions with regard to voices is problematic since a vocal stimulus always involves a temporal element and cannot be entirely static. Even though attempts have been made to overcome this issue, previous studies using brief vocal stimuli have primarily focussed on attractiveness of the speaker, neglecting other potentially important traits (Apicella & Feinberg, 2009; Ferdenzi et al., 2013; Babel et al., 2014). By examining how trustworthiness is perceived in brief, socially relevant speech signals, the current study expands on this approach.

Trait dimensions

Research on face perception has also shown that numerous personality traits can be reduced to summary dimensions such as aggressiveness, dominance and trustworthiness and that face identities can be represented as locations in a multidimensional space (Valentine, 1991; Todorov et al., 2008). Although various labels have been proposed for these dimensions (Fiske et al., 2007; Oosterhof & Todorov, 2008; Sutherland et al., 2013; Scherer, 1972; Zuckerman & Driver, 1989), there seems to be a common understanding that traits can be summarised in a two-

dimensional space, with one dimension emphasizing warmth/likability/trust and the other emphasizing strength/dominance.
In a recent study, McAleer et al. (2014) have investigated the possibility of similar dimensions and evaluative spaces in voice perception. This research will be considered in some detail as it is highly relevant within the current context.
In an online voice rating experiment participants were asked to rate sub-second vocal utterances of the word 'hello' on one of 10 personality traits. The results showed that personality judgements of brief utterances from unfamiliar speakers are consistent across listeners and that a two-dimensional 'social voice space', adequately summarises ratings in both male and female voices. Similar to the dimensional space for personality perceptions from faces, the dimensions emphasize valence (i.e. trust, likeability) and dominance, each driven by differing combinations of vocal acoustics.

Regarding the underlying acoustical information of the valence dimension, the authors found that, for female voices, harmonics-to-noise ratio (HNR), glide and intonation were involved in explaining valence: a more positive perceived valence was associated with a rising intonation whereas a more negative valence was associated with a falling intonation. Since, particularly for female voices, the valence dimension is closely related to trust, manipulating voice features that are involved in explaining valence can, in turn, be assumed to influence perceived trustworthiness.
In male voices, HNR and pitch were the main determinants of perceived valence: an average higher pitch is associated with increased valence, possibly because it brings the pitch closer to that of females, resulting in an increase in positive traits due to stereotyping (Ohala, 1984). There are some inconsistencies in the results as, in male voices, valence is explained by an average higher pitch and increases as a voice is perceived as more dominant. Yet, they also found that lower pitched male voices were perceived as more dominant than higher pitched ones. Notably, more than half of the variance was explained by factors other than pitch and HNR, which were not specified by the authors, indicating a more complex picture of how personality impressions are derived from voices.

Objectives and rationale of the present study

The present study builds upon the findings by McAleer et al. (2014). Since, in their study, a subsample of only 28 participants rated voices for their trustworthiness, a larger sample is needed to more precisely estimate effects. Additionally, by investigating more closely how trustworthiness is perceived in voices, it is attempted to contribute to a more straightforward conception and help parse out the relations between the traits. Although participants were only tested on perceived trustworthiness, i.e. the subjectively estimated degree of another person's trustworthiness, rather than the accuracy of the judgements or the actual trusting behaviour in interaction, the implications extend beyond personality judgements.
After all, our social interactions are influenced by judgements of personality. Due to its evolutionary link to cooperation and survival, trust is a central aspect of human communication, the foundation for functioning relationships, interpersonal interaction, effective collaboration (Tsankova et al., 2013) as well as the cooperation and mutual adaptation needed to surmount problems (McAllister, 1995; Thompson, 1967).
According to ecological theory and the overgeneralization hypothesis, we infer enduring attributes from momentary states or characteristics (Zebrowitz & Collins, 1997), allowing for rapid judgements of personality in order to establish the intent of

another person. Although these judgments may not necessarily be accurate, they help us to estimate whether or not the person can safely be approached (Oosterhof & Todorov, 2008; McArthur & Baron, 1983) and to adapt our behaviour accordingly, in line with a Gibsonian tenet that 'perceiving is for doing' (Gibson, 1979; Fiske, 1992; Gangestad et al., 1992). Such generalizations from snapshot images to enduring attributes appear to hold true for first impressions from faces (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008; Verosky & Todorov, 2010; Said et al., 2009; Zebrowitz et al., 2010). Likewise, it has been suggested that, from brief vocal stimuli we readily form impressions of attractiveness (Apicella & Feinberg, 2009; Ferdenzi et al., 2013; Babel et al., 2014), which has recently been extended to other traits, including trustworthiness (McAleer et al., 2014).

Additionally, at the neural level, it has been shown that we make judgements of identity and attractiveness based on stored prototypes (Latinus et al., 2013; Langlois, & Roggman, 1990; Leopold et al., 2001; Latinus & Belin, 2011; Bestelmeyer et al., 2011), which are explained by acoustical variables related to sexual dimorphism in humans, such as fundamental frequency (f0) and formant dispersion (Latinus et al., 2013). Formant dispersion, defined by the averaged difference between successive formant frequencies, is tied to both vocal tract length and body size (Fitch, 1997) and F0 refers to the rate of vibration of the vocal folds as the primary acoustic property perceived as pitch: male vocal folds are typically longer and vibrate at a lower rate, providing a lower-pitched voice, whereas female vocal folds tend to be shorter and vibrate at a higher rate, leading to a generally higher pitch (e.g. Ohala, 1984).
In the present study, it is explored whether judgments of personality, specifically of trustworthiness, also rely on comparison to a prototype similar to the one used to establish identity. If this is the case, the underlying acoustical properties could be identified, opening up a wide range of potential applications – both in everyday life and interactive technology.

Potential applications

In the present era of increasing computerisation and technological advances, speech synthesis and computer-generated speech are hugely promising areas for applications. For example, as competition increases in business-to-consumer e-commerce, online retailers need to attract and retain more customers and the importance of real-time human-to-human contact in establishing and maintaining relationships with customers has been increasingly respected by online merchants (Qiu & Benbasat, 2005). By incorporating elements of a trustworthy voice into automated voice messages, live help and computer-generated text-to-speech services, the customer service sector could be improved.

Other settings using natural speech include, but are not limited to, telephone interviews. A great majority of refusals occur within the opening phrases of an interview, suggesting that respondents form an attitude towards both the subject and interviewer early on in the conversation and accept or refuse to participate accordingly (Oksenberg et al., 1986). Identifying the features of a trustworthy voice would be of advantage since the interviewer must establish trust and create a sufficiently positive impression with only vocal cues and within the opening phrase.
Although, in that particular study, the trustworthiness scale was eliminated during the analyses since all test voices were rated uniformly high, it may still play a vital role in

interview settings. Furthermore, the results may have been influenced by the effect of vocal characteristics on personal attributions which may depend on the sex of both the speaker and the listener.

Potential mediating factors

Sex of Participants.

Social role theory. It has been shown that females are more accurate on personality judgements (Letzring, 2010; Vogt & Colvin, 2003) and reliably outperform males in judging nonverbal behaviour (Hall, 1984), possibly due to a tendency to be more interpersonally oriented and attuned to social cues (Cross & Madson, 1997; Moskowitz et al., 1994). In a study on the differences between male and female ratings in the perception of the Big Five factors at various levels of acquaintance, Winquist et al. (1998) also found a female positivity effect - a consistent trend indicating that women assign more positive trait ratings to others than men. According to social-role theory (Eagly, 1987), the social roles men and women occupy dictate different gender-based expectancies, skills and beliefs that lead to different behaviour. Typically, the female gender role fosters communal qualities (i.e. interpersonally oriented and socially sensitive), whereas the male role can be described as agentic, dominant, competitive and task-oriented (Bem, 1983; Eagly, 1987, 1995; Siann, 1994). Hence, since gender roles differentially inform how males and females behave, they might also affect how men and women perceive others.

Sex differences in personality impressions from voices. Previous research taking the sex of participants into account has primarily focused on the perception of attractiveness in voices (e.g. Babel et al., 2014). In support of this approach, McAleer et al. (2014) found that ratings were largely consistent across both male and female participants, with the exception of attractiveness.

Ecological theory and the relevance of attractiveness. According to McAleer et al. (2014), the valence dimension, of which trustworthiness is a main component, has the strongest effect in explaining female vocal attractiveness and is strongly, yet marginally weaker than dominance, correlated with male vocal attractiveness (McAleer et al., 2014). Also, it is known that averaging both faces and voices can increase attractiveness (Bruckert et al., 2010; Jones et al., 2007a; Langlois & Roggman, 1990), which, in turn, can increase trustworthiness (Little et al., 2012; Little et al., 2013). In light of the overgeneralization hypothesis, people with more attractive faces are judged more positively on a host of dimensions, such as social competence, intelligence and health (Eagly et al. 1991; Feingold, 1992; Langlois et al., 2000; Zebrowitz et al., 2002; Zebrowitz & Rhodes, 2004). At a behavioural level, these impressions are accompanied by preferential treatment of attractive people in many domains, including occupational settings and the judicial system (Langlois et al., 2000; Zebrowitz, 1997).
This halo effect suggests that attractive people would also be perceived as more trustworthy (e.g. Darby & Jeffers, 1988; DeSantis & Kayson, 1997). Furthermore, the ecological perspective intersects with evolutionary theories, propposing that trustworthiness plays a role in mate selection and is a desirable trait in a long-term partner (Vukovic et al., 2010; Buss, 1999; Fletcher, 2002), which is confirmed by the finding that other characteristics, such as health, status and resources, are rated in a

more variable fashion across studies but are almost always ranked below trustworthiness (Li et al., 2002).

Underlying acoustical information. The general consensus within English is that, for female voices, a slightly higher-than-average pitch is considered more attractive, whereas a slightly lower-than-average voice is more attractive in male speakers (Tuomi & Fisher, 1979; Apple et al., 1979; Zuckerman & Miyake, 1993; Saxton et al., 2006; Riding et al., 2006; Puts et al., 2006; Feinberg et al., 2008). In contrast, it has also been suggested that lower pitched voices are stereotypically associated with threat and dominance, whereas higher pitched voices imply nonthreatening attributes such as submission, appeasement and friendliness (Ohala, 1984). Collectively, research findings suggest that, in males, masculine characteristics (such as a low voice) are associated with some traits that are desirable in a mate and other traits that are not desirable in a mate (Gangestad & Simpson, 2000; Fink & Penton-Voak, 2002; Little et al., 2002).
There is also evidence that men demonstrate preferences for feminine characteristics in women's faces (Jones et al., 2007b; Perrett et al., 1998; Welling et al., 2008) and voices (Collins & Missing, 2003; Feinberg et al., 2008; Jones et al., 2008; Jones et al., 2010a) as these possibly serve as cues for their average estrogen level (Abitbol et al., 1999; Feinberg et al., 2006b; Law Smith et al., 2006).
Recent research on the perception of attractiveness in brief vocal stimuli has also shown that, while ratings by both sexes were highly correlated, men generally rated fellow males as less attractive than women did, but both men and women had similar ratings of female voices (Babel et al., 2014).

Overall, it seems reasonable to infer from social-role and ecological/evolutionary theories as well as studies on impressions of attractiveness that males and females will perceive same- and opposite-sex voices differently. Due to a lack of research on the perception of trustworthiness in voices, the possibility of such an effect will be taken into account and the nature of it, if any, will be explored.

Age of participants. Another potentially relevant factor that has been disregarded in existing research within the given domain is related to age differences between participants. Again, research on face processing has shown that, across the adult age, a positivity bias evolves as age contributes a significant variance in the judgment of neutral and negative faces: with age, people judge them less negatively (Charles et al., 2003; Czerwon et al., 2011). According to socioemotional selectivity theory (Carstensen, 1993, 1995; Carstensen & Charles, 1998; Carstensen et al., 1999), people place more value on positive and emotionally meaningful goals as they get older, investing more cognitive and behavioural resources in obtaining them. This greater emphasis on emotional goals promotes emotion regulation, the maintenance of positive affect and the decrease of negative affect (Carstensen et al., 2000).
If similarities between face and voice processing hold with regard to age related differences, it can be inferred that, in the current study, older participants will judge neutral and untrustworthy voices less negatively than younger participants.

Adaptation effects. Another factor that might be of relevance to the present study concerns adaptation effects, i.e. the phenomenon that exposure to faces or voices biases perceptions of subsequent stimuli (Leopold et al., 2001; Leopold et al., 2005; Little et al., 2005; Little et al., 2008; Rhodes et al., 2004; Webster et al. 2004, Little et al. 2013; Webster & Macleod, 2011; Bestelmeyer et al., 2010).

Audition is susceptible to such effects and adaptation occurs for aspects of sound such as pitch and loudness (D'Alessandro & Norwich 2009; Phillips et al., 2007) as well as for human voice identity (Belin & Zatorre, 2003).
In face research, it has been demonstrated that opposite-sex aftereffects can be induced. For example, exposure to female faces can cause subsequent ambiguous faces to appear more male (Webster et al., 2004; Bestelmeyer et al., 2008; Jaquet & Rhodes, 2008; Little et al., 2005). Likewise, and yet again underlining the similarities between face and voice processing, exposure to voices of one gender causes subsequent voices to be perceived as being of the opposite gender (Schweinberger, et al., 2008).
In the current study, the main interest is not the relationship between the speech signals, but the perceived trustworthiness of different and independent vocal stimuli. By providing an untimed break between the blocks of male and female voices as well as randomizing the stimuli within each block, adaptation effects are intended to be avoided. Moreover, if adaptation occurs nonetheless, it will be detected as participants are assigned to different rating groups, allowing for an intergroup comparison of the results.

Purpose of the present study

The aim of this experiment is to validate the findings by McAleer et al. (2014), with a particular focus on how trustworthiness is perceived in voices and conveyed by novel speakers. Akin to a first impression or 'zero acquaintance' scenario, the design is based on a single word ('hello') and no contextual information is provided.
Using advanced morphing technology, an approach that has been found to provide useful methods for bringing out phenomena in the auditory domain (Schweinberger et al., 2008; Bestelmeyer et al., 2010; Latinus & Belin, 2011), morph continua between voices previously rated high and low on trustworthiness (McAleer et al., 2014) were created for both male and female voices. Thus, full control of the properties of the stimuli could be established to precisely examine whether manipulation towards an averaged trustworthy voice would shift perception of it towards trustworthiness and whether exposure to untrustworthy stimuli would shift perception in the opposite direction. By creating vocal caricatures of the prototypical voices it could also be tested whether acoustically exaggerated versions would enhance the effect.

It is hypothesized that, for both male and female voices, there will be a positive linear relationship between the morph steps and perceived trustworthiness. As voices are manipulated towards a trustworthiness/untrustworthiness caricature, trustworthiness ratings are expected to be higher/lower, respectively.
In addition to possible adaptation effects, the mediating factors taken into account are:

- The sex of participants: It will be explored whether ratings by females differ from ratings by males.
- The age of participants: It will be examined whether older participants rate neutral and untrustworthy voices as more trustworthy sounding than younger participants.

Methods

Ethics Statement

Experimental procedures were approved by the University of Glasgow ethics committee and the study was conducted in accordance with the ethical standards laid down in the Declaration of Helsinki (World Medical Association, 1964).

Participants gave informed consent prior to the experiment, via reading a series of statements regarding freedom to withdraw, anonymity of responses and secured storage of data (Appendix A). As the experiment was carried out online, they confirmed that they had read and agreed to these statements by registering on the website. Participants were not permitted to take part without providing consent.

Design

This parallel-group mixed-model experiment, with the degree of voice manipulation as a within-subjects factor and the rating group to which participants were assigned as well as the sex of participants as between-subjects factors, examined whether voices manipulated towards a trustworthiness caricature would be perceived as more trustworthy sounding than those manipulated towards an untrustworthiness caricature.
To control for potential carryover or order effects, participants in group FM experienced a block of randomized female voices first, followed by male voices, whereas for participants in group MF this order was reversed. Yet, all participants were exposed to the same stimuli and the groups were treated equally in the statistical analyses.

Participants

Four hundred twenty-two participants took part in this voice rating experiment (297 female, 33.5 $\pm$ 17.5 years; 125 male, 30.5 $\pm$ 13.5 years). Participants were recruited via email, social networking sites and word-of-mouth advertising and directed to the web address of the experiment.

Stimuli

Eighteen different voice stimuli were used (9 male/ 9 female). None of the speech signals were taken from actual people, but they were made from a continuum of morphs between a stereotypical untrustworthy and a stereotypical trustworthy voice.

The synthetic voices were produced using STRAIGHT, a speech analysis, modification and synthesis system for manipulating voice quality, timbre and pitch (Banno et al., 2007), in Matlab (version 7).

In a previous study (McAleer et al., 2014), the word ‘hello’ was extracted from voice samples of 64 speakers (all Scottish; 32 male; 28.5$\pm$ 10.2 years) reading out a passage of text. These voices were, amongst other traits, rated on perceived

trustworthiness and provided the basis for the generation of the stimuli. For both male and female voices, the synthetic voices used in the present experiment correspond to 9 morph steps on a continuum from very untrustworthy to very trustworthy, whereby the 3rd and 7th step were prototypical untrustworthy/trustworthy voices, created by averaging together the bottom and top 25% of the voices previously rated low and high on trustworthiness, respectively. The 5th step was an overall average of the voices and, nullifying any effect, is considered to be neutral. The 1st and 9th steps were caricatures of the prototypical un-/trustworthy voices and, the distance to the mean being equal, generated by exaggerating the features associated with low or high ratings of perceived trustworthiness, respectively. Morph steps 2 and 8 as well as 4 and 6 were steps in-between and, though in opposite directions, equal in their distances to the mean and adjacent steps.

Stimuli had an average duration of 391ms ± 65.1ms for male and 390ms ± 64.1ms for female voices. 'Hello' was selected because it is a brief, familiar and socially relevant word, with a medium-to-high range of common usage (British National Corpus).

Procedure

Participants completed the experiment between July and December 2013. As the experiment took place online, control over listening environment was limited. Yet, it has been shown that data from online experiments is comparable to data from lab-based experiments (Germine et al., 2012, Horton et al., 2011).

Participants were blocked by gender and pseudo-randomly assigned to one of two groups, counterbalanced across subjects. Depending on the group, they were either presented with a block of 9 male voices first, followed by a block of 9 female voices or vice versa.

At the start of each block, participants were asked, "For each voice, please rate how trustworthy it sounds, that is, how much you would be ready to trust that person."
No contextual information for the experiment was given.
They indicated their response on a visual analogue scale (VAS), ranging from very untrustworthy to very trustworthy. By using a continuous rating scale some of the pitfalls of discrete rating scales (e.g. Likert scales) could be avoided, such as the limited number of choices and the implication that the space between each possibility is equidistant. There is also evidence that the metrical characteristics of VAS are superior to those of discrete scales, leading to more accuracy in the results as a wider range of statistical methods can be applied (Reips & Funke, 2008).
The voices were blocked by gender and each voice was heard once per discrete block with presentation order randomised in both blocks to avoid possible order or carryover effects.

Participants were given the opportunity to take an untimed break between the blocks.

Results

Data analysis

The data was analyzed using SPSS (IBM Corp., version 21).
For the data to be included in the analysis participants had to complete the experiment and respond to each stimulus, not omitting any. Based on this criterion, none of the data was excluded.
Data preprocessing involved the conversion into a standardized format, i.e. z-transforming the data and adjusting the individual scores accordingly, enabling more efficient procedures as less modification is required in order to perform statistical analyses.

Grouped by voice gender, it was tested whether there was an effect of voice manipulation by conducting repeated measures ANOVAs with morph step (i.e. degree of voice manipulation) as a within-subjects factor (9 levels - 1 and 9: caricatures of un-/trustworthy voices, respectively; 2–8: equidistant steps in-between), controlling for the group to which participants were assigned (2 levels, FM/MF – experienced female/male voices first, respectively) and participant gender (2 levels: male/female) as between-subjects factors.
Due to considerably lower numbers of male raters (group MF: n_{male}=63/ n_{female}=149; group FM: n_{male}=62/n_{female}=148), a direct comparison between the results of male and female participants should be viewed tentatively. Yet, any effect of the sex of participants will be considered in the analysis and examined qualitatively as indicative of trends and tendencies.
A potential between-subjects effect of participant age was analyzed separately.

Since the within-subjects factor has more than 2 levels, the assumption of sphericity had to be tested and Mauchly's test indicated that, for both male and female voices, it had been violated (*approx. chi-square* = 760.54, $p < .001$/*approx. chi-square* = 368.745, $p < .001$, respectively).
To avoid an increased risk of a Type I error and a positively biased F-statistic, degrees of freedom were corrected using Huyn-Feldt estimates of sphericity for female voices (*epsilon* = .804) and Greenhouse-Geisser estimates for male voices (*epsilon* = .557).

Results of the Analyses

For both male and female voices, it was hypothesized that there would be an effect of morph step (i.e. the degree of voice manipulation).

Means and standard deviations of the voice ratings are provided in Table 1. As can be inferred, mean ratings of female voices are all fairly high and no particular pattern becomes apparent. For male voice, on the other hand, mean responses increase as voices are manipulated towards a trustworthiness caricature (morph step 9).

Table 1
Means and standard deviations of ratings for female and male voices

	FEMALE VOICES		MALE VOICES	
Morph step	Mean	Standard deviation	Mean	Standard deviation
1	287.978	100.882	189.775	101.077
2	291.216	90.114	201.687	98.736
3	281.308	91.021	212.624	93.848
4	289.147	90.554	221.431	88.813
5	283.120	90.335	240.382	89.066
6	287.363	87.419	264.960	90.897
7	275.302	94.614	282.322	94.392
8	280.503	93.636	289.771	96.166
9	273.219	99.122	309.428	99.142

The prediction that there would be an effect of morph step was tested by conducting repeated measures ANOVAs as outlined above.
The ANOVA results show that there is a main effect of morph step for both male and female voices, but it is stronger and more significant for male voices (male voices: $F(4.458, 1863.439) = 110.64$; $p < .001$; female voices: $F(6.429, 2687.521) = 2.967$; $p = .006$).
Figure 1 illustrates that, as male voices were manipulated towards un-/trustworthiness, they were perceived as less/more trustworthy sounding, respectively. Whereas this relationship appears to be linear, there is no recognizable pattern for female voices as all the ratings were, on average, fairly high. The graph is even suggestive of a weak negative, though not linear, relationship between manipulation towards a trustworthiness caricature and responses.

A morph step x group interaction that was found for male voices ($F(4.458, 1863.439) = 5.489$; $p < .001$) is illustrated in Figure 2. On average, voices manipulated towards untrustworthiness were perceived as less and those manipulated towards trustworthiness as more trustworthy sounding in the group presented with female voices first than in the group exposed to male voices first.
There was no such interaction for female voices ($F(6.429, 2687.521) = 1.129$; $p = .342$), but a group x participant sex interaction ($F(1, 418) = 5.319$; $p = .022$), as can be seen in Figure 2. On average, after a block of male voices, females rated female voices as more trustworthy sounding than males. This interaction is also displayed in Figure 3 and is, compared to the morph step x group interaction for male voices ratings, not as strong and less significant.

Figure 3 also illustrates a main effect of participant sex for male voices ($F(1, 418) = 2.059$; $p = .009$). Largely, women perceived them as more trustworthy sounding than men. This effect was stronger in the rating group exposed to male voices first, although the overall effect of voice manipulation was, for both male and female participants, weaker in that group.
In the group presented with female voices first, ratings by male and female participants were similar, with the exception of morph step 8 and 9 (i.e. caricatures of trustworthy voices). The morph step x group interaction was greater for male subjects

as, after a block of male voices, they rated the voices as less trustworthy sounding than female participants.

For male voices, no significant effects were found for group (F(1, 418) = 6.908; p = .497), a morph step x sex of participant interaction (F(4.458, 1863.439) = 1.768; p = .125), a group x sex of participant interaction (F(1, 418) = 2.059; p = .152) and a potential 3- way interaction of morph step x group x participant sex (F(4.458, 1863.439) = .425; p = .811).
For female voices, no significant effects were found for group (F(1, 418) = .396; p = .529), sex of participant (F(1, 418) = .023; p = .881), a morph step x group interaction (F(6.429, 2687.521) =1.129; p = .342), a morph step x sex of participant interaction (F(6.429, 2687.521) = 1.778; p = .094) and a potential 3-way interaction (F(6.429, 2687.521) = 1.170; p = .318).

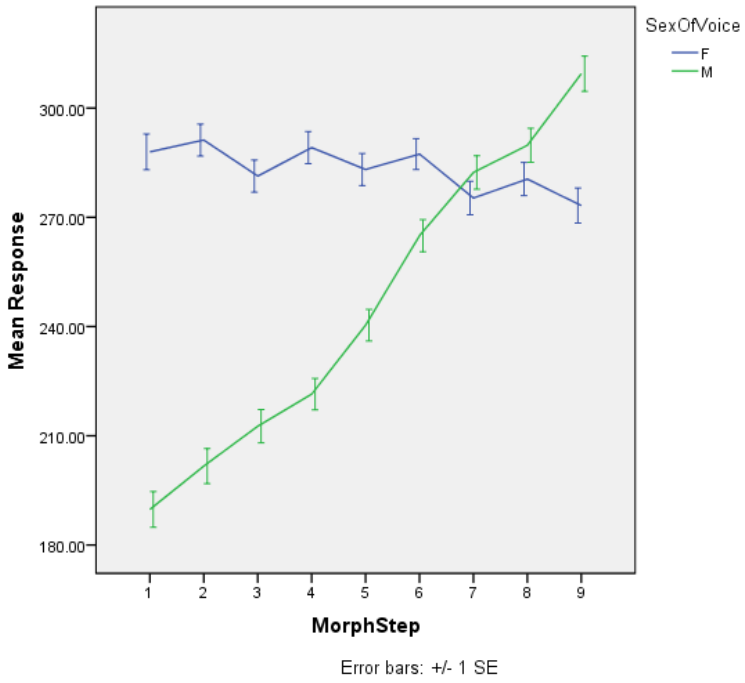

Figure 1. Mean responses as a function of morph step for both male (M) and female (F) voices.

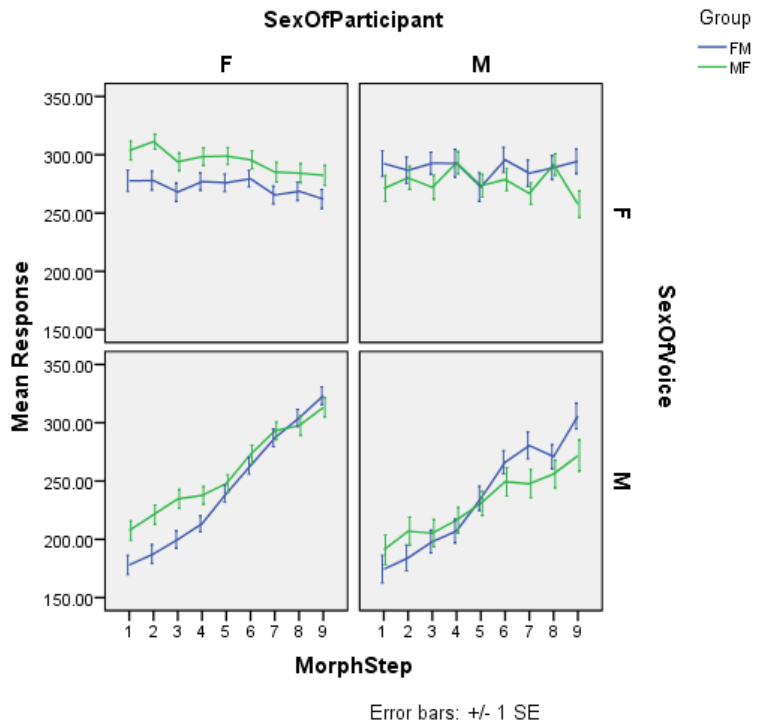

Figure 2. The effect of rating group across male (M)/female (F) voices and participants (Groups MF/FM were presented with male/female voices first, respectively).

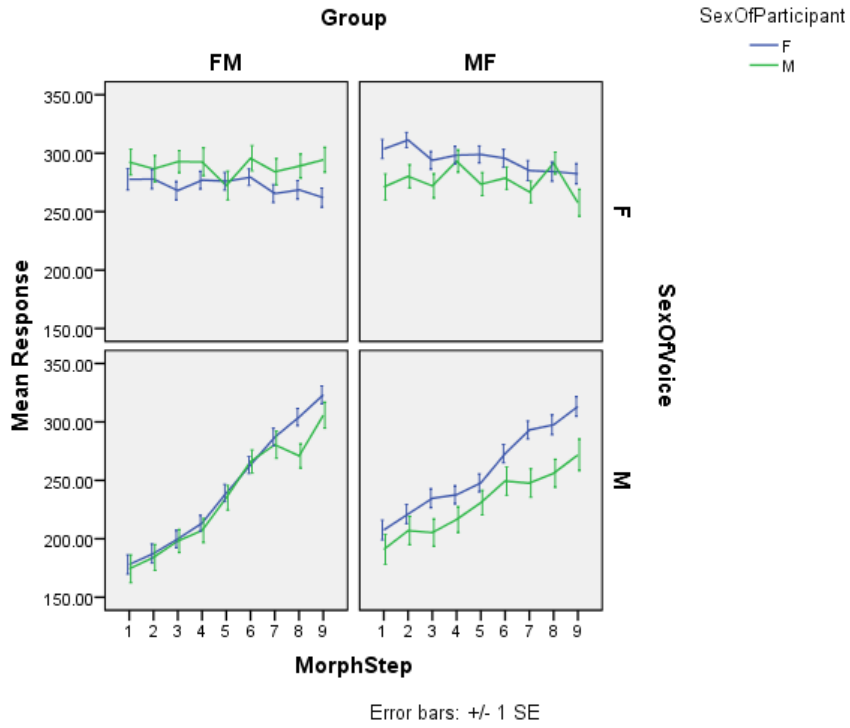

Figure 3. The effect of sex of participant (M/F) across groups (MF/FM: presented with male/female voices first, respectively) and both male (M)/female (F) voices.

For the purposes of the present study the effect of morph step was of main interest and significant for both male and female voices. A positive linear relationship

between the degree of manipulation towards trustworthiness and voice ratings was predicted and tested by correlating responses and morph steps (see Table 2).

Table 2
Results of the correlation analyses for both male and female voices, in total and by rating group

	All	**Group FM (female voices first)**	**Group MF (male voices first)**
Male voices	r= .387 p< .001 n= 3798	r = .469 p < .001 n = 1899	r = .308 p < .001 n = 1899
Female voices	r = -.048 p = .001 n = 3798	r = .030 p = .096 n = 1899	r = .068 p < .002 n = 1899

As can be inferred from Table 2, the results were significant for both male and female voices, with a moderate positive correlation ($r(3796) = .387$, $p < .001$) for male and a weak negative correlation ($r(3796) = -.048$, $p < .001$) for female voices. This is also illustrated in Figures 4 and 5, showing scatterplots of the results with added regression lines.

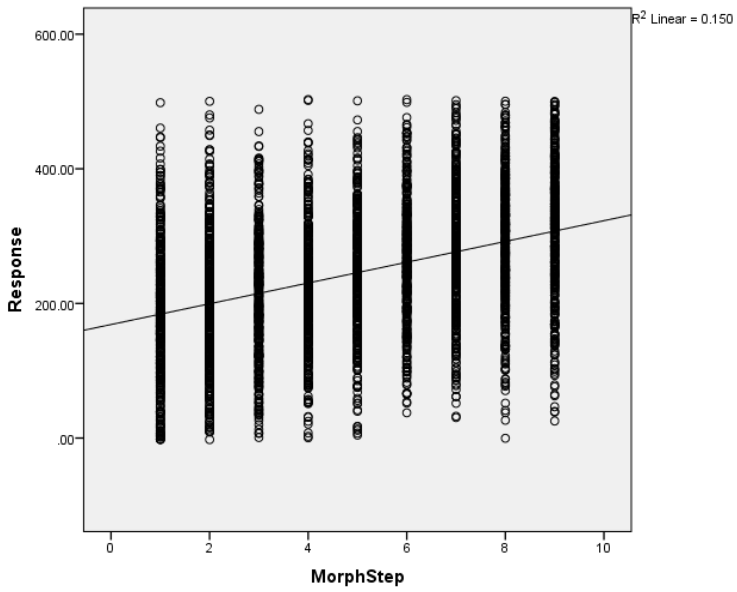

Figure 4. Male voices – Scatterplot with added regression line: Responses as a function of morph step.

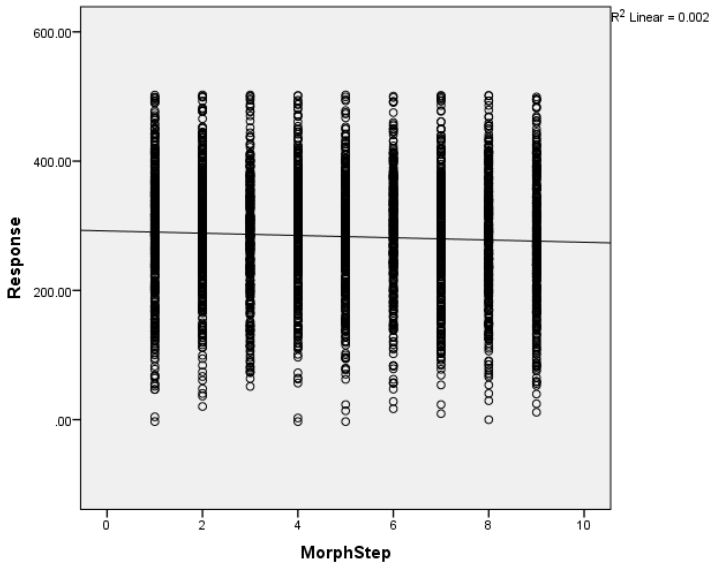

Figure 5. Female voices – Scatterplot with added regression line: Responses as a function of morph step.

Considering the responses altogether, 15% and 2% of the variation in the data can be attributed to the degree of voice manipulation in male and female voices, respectively. By analysing the different cells of the design separately a more complete picture can be obtained (see Table 3).

Table 3
Results of the correlation analyses for male/female participants and voices, separately for both rating groups

		Group FM		**Group MF**	
Sex of Participant	Voice	*Male*	*Female*	*Male*	*Female*
Male		r = .457 p < .001 n = 558	r = .001 p < .498 n = 558	r = .258 p < .001 n = 567	r = -.024 p = .283 n = 567
Female		r = .473 p < .001 n = 1326	r = -.041 p = .131 n = 1341	r = .332 p < .001 n = 1332	r = -.084 p < .001 n = 1332

Only one of the groups rating female voices showed significant results: female participants who experienced male voices first ($r(1330) = -.084$, $p = .001$), explaining no more than 0.7% of the variance (see Figure 8). This weak, yet significant, negative correlation contrasts the predicted positive correlation. Due to the absence of a significant correlation, r^2 could not even be calculated for the remaining groups. Thus, for female voices, there is insufficient evidence in support of the hypothesis.
In contrast, for male voices, positive correlations between the responses and manipulations towards a trustworthiness caricature in all of the four groups support the hypothesis. Though only moderate, these are still highly statistically significant due to the large sample size (Group FM - male participants: $r(556) = .457$, $p < .001$; female participants: $r(1324) = .473$, $p < .001$; Group MF – male participants: $r(565) = .258$, $p < .001$; female participants: $r(1330) = .332$, $p < .001$).

The correlation coefficients also reflect the group effect (see Figure 2). In the group exposed to female voices first there is a stronger correlation between morph step and response. Both male and female participants rated male voices manipulated towards an untrustworthiness caricature as less and those manipulated towards a trustworthiness caricature as more trustworthy sounding if they had experienced female voices first (male participants: $r($556) = .457, p < .001; female participants: r(1324) = .473, p < .001) as compared to the group exposed to male voices first (male participants: r(565) = .258, p < .001; female participants: r(1330) = .332, p < .001). Overall, the differences between the correlation coefficients per group per participant gender show that this effect was stronger for males (r_{FM}–r_{MF} = .199) than it was for females (r_{FM}–r_{MF} = .141), whereas the mean strength of the relationship between response and degree of voice manipulation was, across both groups, greater for females (*mean* $r_{FM,MF}$ = .4025; compare males: *mean* $r_{FM,MF}$ = .3575).

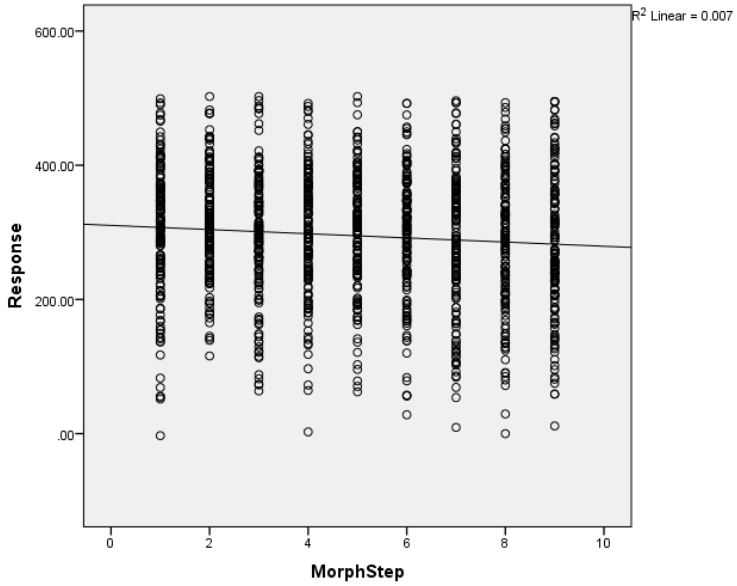

Figure 6. Female voices - Scatterplot with added regression line: Responses by female participants in the group presented with male voices first.

Reliability. To test the internal consistency of the results, measures of inter-rater reliability (Cronbach Alpha) were produced, indicating reliability of judgments across participants within groups. Consistent with the previous analyses and in consideration of the main effects and interactions alpha was calculated separately for each cell in the design (see Table 4).

Table 4
Cronbach Alpha scores for male/female voices and participants per rating group (alpha greater than .85 is considered to be high)

GROUP

	FM		*MF*			*Mean alpha*		
Sex of Voice Participants	Male	Female	Male	Female		Male	Female	All
Male	.818	.891	.874	.877		.846	.884	.865
Female	.877	.921	.850	.924		.8635	.9225	.893
Mean alpha	.8475	.906	.862	.9005				

From Table 4, it can be seen that, overall, the scores are reasonably high for male (all Alphas >.8) and high for female (all Alphas >.85) participants.
Even though, on average, female raters appear to be marginally more consistent in their ratings (females: $mean_{alpha}$ = .891; males: $mean_{alpha}$ = .865), this is not the case across all groups as, for male voices in the group exposed to male voices first, Cronbach alpha is higher for male raters than it is for females (*alpha* = .874 and *alpha* = .850, respectively).
Notably, despite lacking a recognizable pattern, the inter-rater agreement is highest for female voices, across both participant sexes (male participants: *alpha*=.884; female participants: *alpha* = .923) and rating groups (group FM: $mean_{alpha}$ = .906; group MF: $mean_{alpha}$ = .901).

Testing a potential age effect

It was hypothesized that older subjects would be more likely to rate neutral and untrustworthy voices as more trustworthy sounding than younger participants.

Grouping by age. A flaw regarding previous research on age effects in personality perception is a lack of agreement on how participants should be grouped by age. It remains a question for further research to explore whether changes occur gradually or at which age stages they may occur and, accordingly, what age spans would be appropriate for the analyses.
Even though a grouping of participants into subsamples of 'younger', 'middle-aged' and 'older' adults appears to be popular amongst researchers (e.g. Charles et al., 2003; Czerwon et al., 2011), there seems to be no common definition of these age groups.
Hence, in the current study a more pragmatic approach was taken. By establishing equal age ranges, potentially inappropriate labelling could be avoided, yet straightforward statistical analyses could be performed.
Age groups (in years):
1 = 16–20, 2 = 21–25; 3 = 26–30; 4 = 31–35; 5 = 36–40; 6 = 41–45; 7 = 46+

The numbers and proportions of participants in each age group (see Table 5) reveal that there are great differences and imbalances in the distributions.

Table 5
Age distribution: Number and proportion of participants per age group

Age group	*Number of participants*	*Total percentage (in %)*
1 (16 – 20 years)	189	44.79
2 (21 – 25 years)	159	37.68
3 (26 – 30 years)	42	9.95
4 (31 – 35 years)	19	4.5
5 (36 – 40 years)	5	1.18
6 (41 – 45 years)	4	0.95
7 (46+ years)	4	0.95

As can be seen, most of the participants (82.5%) were, at the time of the experiment, 16–25 years old whereas only 8 participants (1.9%) were aged 41+.

As an alternative way of grouping and in an attempt to overcome this imbalance, it had been considered to group participants by establishing equal group sizes. Yet, 4 out of 5 groups would have been within an age range of 16–25 years, whereas only 1 group would have represented all participants aged 25+. Hence, the distributions would have been no less disproportionate and the approach was, therefore, discarded as even more inappropriate for the purpose of examining an age effect.
Due to a lack of homogeneity, to avoid a distortion of the results and to reduce the risk of Type I and Type II errors, a potential age factor was analyzed separately. Again, due to unequal sample sizes, a comparison between the age groups should be viewed cautiously.
For both male and female voices, means and standard deviations per morph step per age group are available in Table A1 (Appendix B).

Data analysis. To test the previously stated hypothesis, repeated measures ANOVAs with morph step as a within-subjects variable (9 levels) and age group as a between-subjects factor (7 levels) were carried out for male and female voices.
Again, Mauchly's test of sphericity indicated that the assumption of sphericity had been violated (male voices: *approx. chi-square* = 786.965, $p < .001$ and female voices: *approx. Chi-square* = 366.532, $p < .001$) and degrees of freedom were corrected using Greenhouse-Geisser estimates for male voices (*epsilon* = .544) and Huynh-Feldt estimates of sphericity for female voices (*epsilon* = .81).

Results of analysis. For male voices, the results show that there is a main effect of morph step ($F(4.349, 1805.008) = 17.422$; $p < .001$), but no significant effect of age group ($F(6, 415) = .882$; $p = .504$) and no morph step x age group interaction ($F(26.096, 1805.008) = .791$; $p = .763$).
Compared to the results from the main ANOVA, it is noteworthy that, whilst the degrees of freedom do not differ greatly and the effect is still significant, the F-statistic is much smaller. Figure 7 illustrates the mean responses of the different age groups as a function of morph step for both male and female voices. From the bottom panel, it can be seen that, even though a linearly ascending trend is recognisable for male voices, it is widely spread between the age groups, suggesting a smaller effect size and illustrating how integrating the variable age group into the model affects the outcome of the analyses.
Considering the results for female voices, this impact becomes even more obvious. After including the factor age group, the effect of morph step, which was small but significant before, is not significant at all ($F(6.482, 2690.092) = .937$, $p = .472$). Furthermore, there is neither an effect of age group ($F(6, 415) = 1.84$, $p = .09$) nor a morph step x age group interaction ($F(38.893, 2690.092) = .904$; $p = .641$). Thus, no pattern becomes recognisable, as can also be seen in Figure 7 (top panel).
It appears from the graphs that participants aged 41–45 show a tendency to judge all female voices and those male voices that are manipulated towards an untrustworthiness caricature as more trustworthy sounding than younger participants, supporting the previously stated hypothesis. In contrast and contradicting the prediction, participants aged 46+ perceived these voices as less trustworthy sounding than younger participants. However, due to the small sample sizes of these age groups (both groups: $n = 4$), the obtained impression is not to be considered representative of the whole populations.

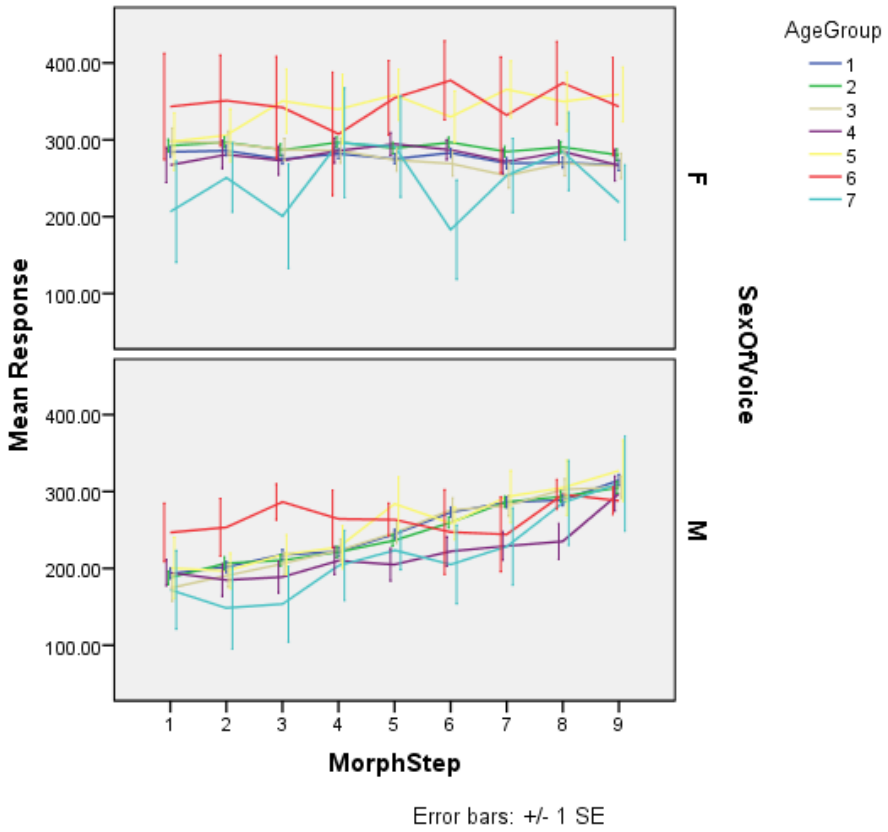

Figure 7. Male (M) and female (F) voices: Mean responses as a function of morph step (separate lines represent age groups, equating to (in years): 1=16-20; 2=21-25; 3=26-30; 4=31-35; 5=36-40; 6=41-45; 7=46+).

Discussion

Review of results in light of the hypotheses and previous research

The aim of the present study was to investigate how trustworthiness is perceived in voices and conveyed by novel speakers.
The online nature of the study is not assumed to have affected the results. Although control of experimental listening environment was limited, it has been shown that data from online experiments is comparable to data from lab-based experiments (Germine et al., 2012; Horton et al., 2011). Additionally, the large sample size and high internal consistency suggest high validity and reliability of the results.

It was examined whether the manipulation of a vocal signal towards an averaged prototypical un-/trustworthy voice would shift perception of it towards un-/trustworthiness and, if so, whether this effect would be enhanced by vocal caricatures of these prototypes.
For male, yet not for female voices, this hypothesis is supported, aligning with research that observers form consistent and reliable impressions from brief exposure to faces (Willis, 2006; Rhodes, 2011; Zebrowitz, 2012) and both extracts of extended speech (Zuckerman, 1989, 1990) and brief vocal signals (McAleer et al., 2014).

For female voices, the statistical analysis showed that, although the degree of voice manipulation seemed to have had an effect, the predicted positive linear relationship

between ratings and manipulation towards a trustworthiness caricature did not become apparent. This may be due to an error in the generation of the stimuli, found after the data had been collected. Thus, the synthetic voices did not represent the initially intended morphing procedures, which may have led to distortions. Hence, the results obtained for female voices, deemed invalid and inconclusive, will not be discussed any further.
For male voices, on the other hand, it can be concluded that listeners show high consistency in their ratings of perceived trustworthiness from brief utterances containing limited information, akin to a first impression. An effect of the degree of manipulation and a moderate, yet highly statistically significant, correlation between responses and voice manipulations demonstrate that the voices were, on average, perceived as more/less trustworthy sounding as they were manipulated towards trustworthiness/untrustworthiness caricatures, respectively.

The hypothesized age effect, assuming older participants to rate neutral and untrustworthy voices as more trustworthy sounding than younger participants, could not be detected. This does not mean that there is no such positivity bias as suggested by socioemotional selectivity theories (e.g. Carstensen, 1993, 1995; Carstensen & Charles, 1998; Carstensen et al., 1999). Rather, the current results are inconclusive:
As participant recruitment was not targeted at specific ages, the distributions of different age groups in the sample were highly disproportionate and imbalanced. A great majority of participants were undergraduate students (i.e. late teens to mid 20's), whereas other age groups were underrepresented, not allowing for a comprehensive comparison.

Underlying acoustical information.
By exaggerating the features specific to the voice prototypes, the type of vocal information used for the evaluation could be identified and both pitch and intonation were involved in explaining trustworthiness.
Considering female voices, the trustworthiness caricature would have been characterized by an average higher pitch and a rise in pitch between the first and second vowel of the word 'hello' whereas the untrustworthiness caricature would have been defined by a lower pitch and slightly falling intonation. Due to the error in the synthesis, the stimuli deviated from this outline, particularly at the lower end of the continuum, representing averaged and caricatured untrustworthy voices.
Nonetheless, compared to previous findings by McAleer et al. (2014), the given variations in vocal properties would reflect their conclusion that valence is strongly associated with trust and a positive perceived valence is characterized by a rising intonation whereas a negative valence is associated with a falling intonation.

For male voices, on the other hand, a lower pitch and rising intonation relate to decreased trustworthiness, whereas the main features of a stereotypical trustworthy voice are a falling intonation and an average higher pitch. Again, this aligns with findings that an average higher pitch is associated with increased valence and, in turn, trustworthiness, implying that a revision of the idea of valence as being strongly positively correlated with dominance, which increases as pitch decreases, would be advisable (McAleer et al., 2014). Harmonics-to-noise ratio (HNR), which was identified as the other main feature involved in explaining the variance in valence, was not analyzed in the current study and the authors do not comment on intonation in relation to male voices, precluding a more comprehensive comparison. Hence, the

decomposition of the stimuli into more specific parameters, including HNR, should provide better insights into the relations between the components.
On the surface, it seems as though intonation might play a role, in alignment with a connection reported between facial features and trustworthiness (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008). Both vocal intonation and facial expression are transient, adjustable and malleable features of their respective modalities, suggesting that the processing of faces and voices operates via equivalent comparisons of the available information to each modality (Belin et al., 2004, 2011, Bruce & Young, 1986; Young & Bruce, 2011).

Comparison to vocal attractiveness. Despite the connection between trustworthiness and attractiveness (Bruckert et al., 2010; Jones et al., 2007; Langlois et al., 1990; Little et al., 2012), the present results differ from previous findings on vocal attractiveness that a lower voice is perceived to be more attractive in male speakers (Saxton et al., 2006; Riding et al., 2006; Tuomi & Fisher, 1979; Apple et al., 1979; Zuckerman & Miyake, 1993; Puts et al., 2006; Feinberg et al., 2008).
In another line of argument, which is supported by the current findings, it has been suggested that masculine characteristics, such as a low voice, are also stereotypically associated with traits that are not desirable, including threat, dominance and dishonesty (Gangestad & Simpson 2000; Fink & Penton-Voak 2002; Little et al., 2002), whereas higher pitched voices imply positive attributes such as friendliness (Ohala, 1984). The current findings suggest that trustworthiness and untrustworthiness may complement these clusters of positive and negative traits, respectively.
The female test voices were, on average, all rated fairly high on perceived trustworthiness and, due to the error, they were all rather high pitched, thus possibly reflecting previous findings that men demonstrate a preference for feminine characteristics in women's voices (Collins & Missing, 2003; Feinberg, et al., 2008b; Jones et al., 2008b; Jones et al., 2010a), which also aligns with a preference for femininity in faces (Jones et al., 2007; Perrett et al., 1998; Welling et al., 2008).

Adaptation effect. It was found that there is a stronger correlation between the degree of voice manipulation and responses for participants in the group exposed to female voices first. They perceived voices manipulated towards an untrustworthiness caricature as less and those manipulated towards a trustworthiness caricature as more trustworthy sounding than participants in the group presented with male voices first. Aligning with previous findings that exposure to faces or voices biases perceptions of subsequent stimuli (e.g. Leopold et al., 2001, 2005; Little et al. 2005, 2008, 2013; Rhodes et al., 2004; Webster & Macleod, 2011) and with studies reporting opposite-sex aftereffects (e.g. Webster et al., 2004, Bestelmeyer et al., 2010), the present results are suggestive of an adaptation effect, which might be explained in light of the underlying acoustical information.
By experiencing female voices first, the listener may have become attuned to a female pitch range, which may have led to a biased perception of the male voices that followed: a very low pitched male voice would create a greater contrast and, thus, be judged as more extreme towards untrustworthiness, whereas a higher voice would bring the pitch closer to that of females and, therefore, be perceived as more trustworthy due to stereotyping (Ohala, 1984).

Sex differences. When comparing ratings by male and female participants, it is important to keep in mind that, due to considerably lower numbers of males in the sample, the results should be viewed tentatively.
Although, on average, responses by females appear to be slightly more consistent, the effect is only marginal compared to previous research, suggesting that women tend to be more interpersonally attuned and oriented than men (Cross & Madson, 1997; Moskowitz et al., 1994). A possible reason might be a difference in intended research outcomes. Previous studies were not concerned with judgments from first impressions or from voices and focussed on the accuracy of personality judgments (Letzring, 2010; Chan et al., 2011; Vogt & Colvin, 2003). In the present study, it was found that participants agree on the features associated with vocal un-/trustworthiness, across both participant sexes and rating groups, supported by high internal consistency. Yet, these judgments may not necessarily be accurate.
In previous studies with research interests similar to the current one, it has either been the case that no effects of participant gender were reported (e.g. Oosterhof & Todorov, 2008) or that no such effects were found except for ratings of attractiveness (McAleer et al., 2014).

Female positivity bias. The present results show that, on average, females rated male voices higher on perceived trustworthiness than male participants, particularly in the rating group exposed to male voices first, which might be a manifestation of the adaptation effect.
This outcome aligns with previous findings that males generally rated fellow males as less attractive than females did (Babel et al., 2014).
In light of evolutionary and mate selection theories (Vukovic et al., 2010; Buss, 1999; Fletcher, 2002), a possible explanation for this effect would be that, at a first and brief encounter, females may be inclined to perceive a voice in terms of the qualities of a potential mating partner. Since trustworthiness is important in mate selection and a desirable trait in a long-term partner (Vukovic et al., 2010; Li et al., 2002) they would be more receptive. In contrast, males, viewing other males as competitors or rivals, may be more suspicious and likely to distrust them.
An alternative explanation revolves around social-role theory (Eagly, 1987), arguing for a differing placement of men and women within a social structure and implying that the expectancies of gender-appropriate behaviour largely determine behaviour. As the female gender role fosters communal qualities, women are stereotypically more socially oriented, striving for both a positive impression and a positive reaction in an interpersonal encounter. The gender-appropriate behaviour would consist of female's positive ratings of others, supporting the notion that gender roles not only impact behaviour differentially for males and females, but also differentially inform how males and females perceive others (Winquist et al., 1998).
However, it should be noted that this effect did not occur in a similar fashion across the rating groups and both male and female voices (possibly due to the error in the generation of female voice stimuli). This and the imbalance in sample sizes of male and female participants suggest that any interpretation of the differences between sexes in the present findings should be viewed tentatively.

Implications

Although, in this study, participants were tested on perceived trustworthiness rather than the accuracy of the judgements or the actual trusting behaviour in interaction, the implications extend beyond personality judgements.

According to the overgeneralization hypothesis, we infer enduring attributes from momentary states or characteristics (Zebrowitz & Collins, 1997), allowing for rapid judgements of personality. A speaker's intent can be established by making an immediate judgment on trustworthiness which, in turn, helps us to estimate whether or not he/she can safely be approached and adjust our behaviour accordingly (McArthur & Baron, 1983; Oosterhof & Todorov, 2008). Such generalizations from snapshot images to enduring attributes have been verified for faces (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008; Verosky & Todorov, 2010; Said et al., 2009, Zebrowitz et al., 2010) and judgments of attractiveness in voices (e.g. Apicella & Feinberg, 2009; Ferdenzi et al., 2013; Babel et al., 2014), which has recently been extended to other traits (McAleer et al., 2014). The current study supports and expands on these findings as it is the first of its kind to focus exclusively on how trustworthiness is conveyed by novel speakers and, though not entirely static, the use of short and socially relevant vocal bursts highlights the similarities with the study of personality judgments from first impressions in faces.

Furthermore, in the design of this experiment the role of the perceiver was emphasized at a behavioural level by asking how much he/she would be ready to trust the speaker rather than just how trustworthy the voice sounds. Thus, the phrasing of the task shifted the position of the participant from a passive recipient towards an active agent.

Potential applications

Underlying potential implications at a behavioural level are clusters of traits that may be associated with the acoustical properties.
For instance, a rise in pitch between the first and second vowel of the word 'hello' may appear like the intonation of a question, suggestive of uncertainty and doubt. In contrast, trustworthiness encompasses attributes such as reliability, confidence, certainty, credibility and honesty – as more likely to be conveyed by a falling intonation, resembling a reassuring statement.
Also, due to its malleable nature, intonation is highly susceptible to manipulation, opening up numerous potential applications. Thus, simply by modifying the way we speak, we could influence how trustworthy others perceive us to be and behave towards us, which could be applied in a variety of settings, from everyday personal life to public, political and judicial settings.

For example, telephone interviewers face the task of establishing trust and creating a positive impression with only vocal cues. Since a great majority of refusals occur within the opening phrases (Oksenberg et al., 1986), a positive attitude towards both the subject and interviewer could be induced early on, resulting in increased response rates. The current results suggest that the initial study on response rates in telephone interviews by Oksenberg et al. (1986) was flawed as insufficient attention had been paid to the definition of the constituents of a trustworthy voice and gender differences in both speakers and perceivers. Hence, it would be necessary to

explore in a real-world setting whether a trustworthy impression could be induced and response rates improved by applying the features associated with trustworthiness to the interviewer's voice.

Furthermore, speech synthesis and computer-generated voices could be addressed as a promising area for applications. The current findings provide a starting point for improvements in the customer service sector. As competition increases, online retailers need to attract and retain customers by implementing new customer support features and real-time human-to-human contact (Qiu & Benbasat, 2005).
Again, in order to attract new customers, trust would need to be established early on and further research is needed to assess the applicability of the current results to computer-generated devices such as automated messages, text-to-speech services and live help in order to improve customer service systems.

Limitations and how to overcome them

Practical issues. Within the current context, some of the issues that arose are of a practical nature and can relatively easily be eliminated.
As has already been pointed out, due to the error in the generation of the stimuli, the results for female voices were inconclusive. Hence, it remains subject to further research to replicate the current study, based on the corrected speech signals. Following from this, it should also be enlightening to investigate possible sex differences in the trustworthiness judgments of female voices and how they compare to the current findings for male voices.
Also, more specific recruitment criteria would lead to a more balanced sample in order to investigate the role of both the sex and age of participants more reliably.
Another aspect that could be taken into account is the cultural background of participants and how it might affect individual personality impressions. Although, at registration for the current experiment, participants provided information on their nationality, this may not necessarily be representative of their cultural influences and was, therefore, not analyzed more rigorously.
From previous research, it is known that there are cultural differences in speech production and perception. For example, pitch differences between males and females are greater in Japanese than in Dutch, possibly due to greater relative extremes in gender stereotypes and expectations (van Bezooijen, 1995), which may lead people to be attuned to different vocal cues. In support of this theory, Babel et al. (2014) concluded that various parameters of acoustic qualities, including local sociophonetic cues and conformity to community speech norms, contribute to the perceived personality of a speaker, indicating a more complex picture of personality judgments from voices.
Even though none of the stimuli used in the present experiment was taken from actual people, the speakers they were based on were all born and raised in Scotland, which may have affected individual perception and judgments.
Additionally, different cultural values may cause people from different cultural backgrounds to be more or less attuned to certain traits. For instance, perceivers from individualistic cultures have been found to be highly attuned to dominance and extraversion, whereas people from collectivist cultures appear to be attuned to conscientiousness and, consequently, more accurate in their judgments of the respective traits (Bond & Forgas, 1984). According to the ecological theory of social perception, we are attuned to adaptively relevant information. Yet, the degree of

relevance may differ across cultures. For instance, trustworthiness may be more relevant in a collectivist culture where people are highly interdependent than in an individualistic one (Zebrowitz & Collins, 1997) where the majority of the participants in the current study came from (as indicated by their nationalities).
Although the high inter-rater agreement suggests that the perception of what constitutes trustworthy and untrustworthy voices is shared cross-culturally, it would be instructive to investigate the role of culture in the perception of trustworthiness systematically, considering factors such as the influence of cultural background on personality judgments or the role of sociophonetic cues.

Methodological issues. The main issue regarding the methodology and design of the current experiment is exemplified by the adaptation effect.
The ultimate aim was to generalize from the results and make predictions about real-world events. Even though zero acquaintance scenarios are part of everyday life and it could be established that we make rapid personality judgments, it is highly unlikely that, in a life-like situation, we would encounter a number of different female voices followed by a number of different male voices or vice versa.
In the present study, the focus was not on the relations between the different stimuli, but, ideally, they would have been judged independently from one another. Yet, the adaptation effect shows that the means employed to achieve this and avoid comparison to previously heard voices, i.e. the randomization of stimuli within each block of male/female voices and the untimed break between the blocks, were insufficient.
The design could, thus, be improved by randomizing the order of all the stimuli and not blocking them by speaker gender. In addition, short distraction tasks could be included between the different stimuli to further ensure that order effects are avoided.

Taking the current results further

There are various ways in which future research might build upon the current findings. In addition to the ones already mentioned, the relations between the features involved in explaining trustworthiness could be parsed out, for instance, by applying a falling intonation to a low pitched male voice or a rising intonation to a high pitched male voice. This would help assessing the validity and generalizability of the current results and provide a method to explore the extent to which the findings are applicable to wider settings.
Furthermore, when considering possible applications it would be advisable to study the role of context effects in personality judgments from brief vocal stimuli.
Even though the importance of contextual information in first impressions of personality from brief exposure to voices has largely been disregarded, research into context effects on impression formation in other domains (e.g. Wyer et al., 1969; Skowronski et al., 1989; Hamilton et al., 1974; Schiller et al., 2009) suggests that the contextual information of interpersonal encounters influences our perception of another person, which can also be assumed to hold true for personality judgments from voices and requires further investigation.

The relevance of contextual information is also linked to another important, yet often neglected, factor in the study of personality judgments: the accuracy of first impressions. After all, the consensus at zero acquaintance found in the current study might reflect shared vocal stereotypes rather than accurate perceptions of people's

trustworthiness. Accuracy can be determined by converging self-ratings and ratings by acquaintances. Yet, this approach has been criticised for showing only modest convergence for a limited number of traits such as dominance and honesty (Zebrowitz & Montepare, 2008; Zebrowitz & Collins, 1997; Olivola & Todorov, 2010). One problem is the misconception that personality is context-independent. Social affordances, i.e. the opportunities for acting, interacting or being acted upon, are inherently connected to a particular social context (Zebrowitz & Collins, 1997) and, although people may accurately infer the momentary state of another person, the impression may not hold when generalised across time and other situations. Thus, to establish how accurate people are in judging the personality of others, a context-based measure of accuracy would be more appropriate (Funder, 2012).

Finally, considering how the current findings may be generalized across wider settings, it also remains a challenge to integrate and reconcile the results across experiments which retain a natural and, therefore, uncontrolled speech signal and those which synthetically manipulate the stimulus. By generating synthetic voices – the approach taken in the present study - the acoustic parameters of the stimuli can be fully controlled, allowing for clear, comprehensive and unambiguous hypothesis testing. Yet, it might also lead to combinations of acoustic-phonetic parameters that would not occur naturally, thus giving listeners tokens which do not approximate natural speech. Hence, the implications of the current study should be neither over- nor underestimated as both approaches are necessary to fully understand the results at hand and assess their applicability.

General Conclusions

Inspired by similarities between face and voice processing, the aim of this study was to investigate how trustworthiness is conveyed by novel speakers using brief vocal stimuli. By synthetically manipulating the speech signal and generating a continuum between stereotypical trustworthy and untrustworthy voices, full control over the acoustical parameters of the stimuli could be established.

It was found that, for male voices, listeners show high agreement in their judgments of trustworthiness from first impressions, in support of the hypothesis that voices manipulated towards a trustworthiness caricature would be rated as more trustworthy sounding, whereas manipulation towards an untrustworthiness caricature would shift perception in the opposite direction. For female voices, the study needs to be replicated as, due to an error in the generation of the stimuli, the results are inconclusive.

This study contributes to the empirical basis for the assessment of personality from speech signals. In establishing the acoustical information involved in explaining the perception of trustworthiness in voices, people and algorithmic devices featuring computer-generated voices may be instructed on alterations to obtain a desired projection, opening up a wide range of potential applications in many areas, including business, advertising and everyday life.

Future research may build upon the current findings by focussing on the generalizability across differing contexts and forms of delivery, integrating and reconciling the current results with studies on natural speech and identifying the role of community-based typicality. It is important that such research avoids essentializing complex speech signals, risking the creation of a misleading understanding of how listeners perceive, categorize and use vocal signals.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice, 13*, 424-446.

Allport, G.W., & Cantril, H. (1934) Judging personality from the voice *Journal of Social Psychology,* 5(1): 37-55.

Apicella C.L., & Feinberg D.R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society B-Biological Sciences*, 276(1659): 1077-1082.]

Apple W., Streeter L.A., & Krauss R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology,* 37: 715–727.

Babel, M., McGuire, G., & King, J. (2014). Towards a More Nuanced View of Vocal Attractiveness. *PLoS ONE* 9(2): e88616. doi:10.1371/journal.pone.0088616

Banno, H., Hata, H., Morise, M., Takahashi, T., Irino, T., & Kawahara, H. (2007). Implementation of realtime STRAIGHT speech manipulation system: Report on its first implementation. *Acoustical Science & Technology*, 28(3): 140-146.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion (Washington, DC),* 6: 269–278.

Belin P., Bestelmeyer P.E., Latinus M., & Watson R. (2011). Understanding voice perception. *British Journal of Psychology,* 102(4): 711-25.
Belin P., Fecteau S., & Bedard C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3): 129-35.)

Belin, P., & Zatorre, R.J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, 14: 2105–2109.

Bem, S.L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs,* 8: 598–616.

Berry D.S. (1990). Vocal Attractiveness and Vocal Babyishness - Effects on Stranger, Self, and Friend Impressions. *Journal of Nonverbal Behavior,* 14(3): 141-53.

Bestelmeyer, P.E.G., Jones, B.C., DeBruine, L.M., Little, A.C., Perrett, D.I., & Schneider, A. (2008). Sex-contingent face aftereffects depend on perceptual category rather than structural encoding. *Cognition*, 107: 353-365.

Bestelmeyer P.E., Latinus M., Bruckert L., Rouger J., Crabbe F., & Belin P. (2011). Implicitly Perceived Vocal Attractiveness Modulates Prefrontal Cortex Activity. *Cerebral Cortex,* 22(6): 1263-1270.

Bestelmeyer, P.E.G., Rouger, J., DeBruine, L.M.,, & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition,* 117: 217–223.

Bezooijen, van, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech,* 38: 253–265.

Bond, M.H., & Forgas, J.P. (1984). Linking person perception to behavior intention across cultures: The role of cultural collectivism. *Journal of Cross-Cultural Psychology,* 15: 337-352.

British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/ (accessed: 26/02/14)

Bruce V., & Young A. (1986). Understanding face recognition. *British journal of psychology,* 77(3): 305-27.

Bruckert L., Bestelmeyer P., Latinus M., Rouger J., Charest I., Rousselet G.A., Kawahara H., & Belin P. (2010). Vocal attractiveness increases by averaging. *Current biology*, 20(2): 116-120.

Buss, D.M. (1998). Sexual strategies theory: Historical origins and current status. *Journal of Sex Research*, 35: 19-31.

Buss, D.M. (1999). *Evolutionary psychology: The new science of the mind*. Boston: Mass., London: Pearson.

Carstensen, L.L. (1993). Motivation for social contact across the life span: A theory of socioemotional selectivity. *Nebraska Symposium on Motivation,* 40: 209-254.

Carstensen, L.L. (1995). Evidence for a life-span theory of socioemotional selectivity. *Current Directions in Psychological Science,* 4: 151-156.

Carstensen, L.L., & Charles, S.T. (1998). Emotion in the Second Half of Life. *Current Directions in Psychological Science,* 7: 144-149.

Carstensen, L.L., Isaacowitz, D.M., & Charles, S.T. (1999). Taking time seriously:A theory of socioemotional selectivity. *American Psychologist,* 54(3): 165–181.

Carstensen, L.L., Pasupathi, M., Mayr, U., & Nesselroade, J.R. (2000). Emotional experience in everyday life across the adult lifespan. *Journal of Personality and Social Psychology,* 79: 644–655.

Chan, M., Rogers, K.H., Parisotto, K.L., & Biesanz, J.C. (2011). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, 45: 117-120.

Charles, S.T., Mather, M., & Carstensen, L.L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology (General),* 132: 310–324.

Cross, S.E., & Madson, L. (1997). Models of the self: self-construals and gender. *Psychological Bulletin,* 122: 5–37.

Czerwon, B., Lüttke, S., & Werheid, K. (2011). Age differences in valence judgments of emotional faces: The influence of personality traits and current mood. *Experimental Aging Research,* 37: 503–515.

Collins, S.A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behavior*, 65: 997-1004.

D'Alessandro, L.M., & Norwich, K.H. (2009). Loudness adaptation measured by the simultaneous dichotic loudness balance technique differs between genders. *Hearing Research*, 247: 122–127.

Darby, B.W., & Jeffers, D. (1988). The effects of defendant and juror attractiveness on simulated courtroom trial decisions. *Social Behaviour and Personality,* 16: 39-50.

DeSantis, A., & Kayson, W.A., (1997). Defendants characteristics of attractiveness, race and sex on sentencing decisions. *Psychological Reports,* 81: 679 – 683.

Elkins, A.C., & Derrick, D.C. (2013). The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. In *Group Decision and Negotiation* (published online: http://ibug.doc.ic.ac.uk/media/uploads/documents/elkins2013sound.pdf, accessed: 03/11/2013)

Eagly, A.H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Erlbaum.

Eagly, A.H. (1995). The science and politics of comparing women and men. *American Psychologist,* 50: 145–158.

Eagly A.H., Ashmore R.D., Makhijani M.G., & Longo L.C. (1991). What Is Beautiful Is Good, but: A Meta analytic Review of Research on the Physical Attractiveness Stereotype. *Psychological Bulletin*, 110(1): 109–128.

Feinberg D. R., DeBruine L. M., Jones B. C., Little A. C., O'Connor J. J. M., & Tigue C. C. (2012). Women's self-perceived health and attractiveness predict their male vocal masculinity preferences in different directions across short- and long-term relationship contexts. *Behavioral Ecology and Sociobiology*, 66(3): 413-8.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The relative role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, 37: 615-623.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G., & Perrett, D. I. (2006). Menstrual cycle, trait estrogen level, and masculinity in the human voice. *Hormones and Behavior*, 49: 215-222.

Feingold A. (1992). Good-looking people are not what we think. *Psychological Bulletin,* 111: 304–341.

Ferdenzi C., Patel S., Mehu-Blantar I., Khidasheli M., Sander D., & Delplanque S. (2013). Voice attractiveness: Influence of stimulus duration and type. Behavior research methods, 45(2): 405-13.

Fink, B., & Penton-Voak, I. S. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science,* 11: 154–158.

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, 63: 877-889.

Fiske S. T., Cuddy A. J., Glick P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences,* 11(2): 77-83.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America* 102: 1213–1222.

Fletcher, G. J. O. (2002). *The new science of intimate relationships*. UK, Cambridge: Blackwell.

Funder D. C. (2012). Accurate Personality Judgment. *Current Directions in Psychological Science,* 21(3): 177-82.

Gangestad S. W., & Simpson J. A. (2000). The evolution of human mating: trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, 23: 573–644.

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: Examination of a functional hypothesis. *Journal of Personality and Social Psychology*, 62: 688-698.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C.F., Chatterjee, G., & Wilmer J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review,* 19(5): 847-57.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, Mass.: Houghton Mifflin; London: Erlbaum.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Baltimore, MD: Johns Hopkins University Press.

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: Changes in connotative meaning. *Journal of Personality and Social Psychology,* 29(5): 649-654.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics,* 14(3): 399-425.

Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior,* 25(5): 295-304.

Jaquet, E., & Rhodes, G. (2008). Face aftereffects indicate dissociable, but not distinct, coding of male and female faces. *Journal of Experimental Psychology: Human Perception and Performance*, 34: 101–112.

Jones, B. C., DeBruine, L. M., Little, A. C. (2007a). The role of symmetry in attraction to average faces. *Perception & Psychophysics*, 69(8): 1273-7.

Jones, B. C., DeBruine, L. M., Little, A. C., Conway, C. A., Welling, L. L. M., & Smith, F. (2007b). Sensation seeking and men's face preferences. *Evolution and Human Behavior*, 28: 439-446.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2008). Integrating cues of social interest and voice pitch in men's preferences for women's voices. *Biology Letters*, 4: 192-194.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour,* 79: 57-62*.*

Kramer, E. (1964). Personality stereotypes in voice: A reconsideration of the data. *Journal of Social Psychology*, 82: 247-251.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin,* 126: 390–423.

Langlois, J. H., & Roggman, L. A. (1990). Attractive Faces Are Only Average. *Psychological Science*, 1(2): 115-21.

Latinus, M., & Belin, P. (2011), Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2: 175. (Published online: doi: 10.3389/fpsyg.2011.00175)

Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin P. (2013). Norm-Based Coding of Voice Identity in Human Auditory Cortex. *Current Biology,* 23(12): 1075-80.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, [Clinical Trial Research Support, Non-U.S. Gov't],* 4(1): 89-94.

Leopold, D. A., Rhodes, G., Muller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of London, Series B*, 272: 897-904.

Letzring, T. D. (2010). The effects of judge-target gender and ethnicity similarity on the accuracy of personality judgments. *Social Psychology,* 41(1): 42 – 51.

Little, A. C., Debruine, L. M., & Jones, B. C. (2013). Sex Differences in Attraction to Familiar and Unfamiliar Opposite-Sex Faces: Men Prefer Novelty and Women Prefer Familiarity. *Archives of Sexual Behavior.*

Little, A. C., Roberts, S. C., Jones B. C., & DeBruine L. M. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *Quarterly Journal of Experimental Psychology*, 65(10): 2018-32.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., . . . , & Hillier, S.G. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B*, 273: 135-140.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sexcontingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society B: Biological Sciences*, 272: 2283–2287.

Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008). Category contingent aftereffects for faces of different races, ages and species. *Cognition*, 106: 1537–1547.

Little, A. C., Feinberg, D. R., DeBruine, L. M., & Jones, B.C. (2013). Adaptation to Faces and Voices: Unimodal, Cross-Modal, and Sex-Specific Effects. *Psychological Science OnlineFirst, XX*(X): 1–9.

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proceedings of the Royal Society B: Biological Sciences*, 269: 1095–1103.

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeier, J. A. W. (2002). The necessities and luxuries of mate preferences: Testing the trade-offs. *Journal of Personality and Social Psychology*, 82: 947-955.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices*. [In press].*

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal,* 38(1): 24–59.

McArthur, L. Z., & Baron, R. M. (1983). Toward an Ecological Theory of Social-Perception. *Psychological Review,* 90(3): 215-38.

Montepare J.M., & Zebrowitz-McArthur L.A. (1987). Perceptions of Adults with Child-Like Voices in 2 Cultures. Journal of Experimental Social Psychology, 23(4): 331-49.

Moskowitz, D. S., Suh, E. J., & Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *Journal of Personality and Social Psychology*, 66: 753–761.

Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica,* 41(1): 1-16.

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986) Interviewer's voices and refusal rates in telephone surveys. *The Public Opinion Quarterly,* 50(1): 97 – 111.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences*. Journal of Experimental Social Psychology*, 46(2): 315-24.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America,* 105(32): 11087-92.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D.M., . . . & Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, 394: 884-887.

Phillips, D. P., Scovil, S. J., Carmichael, M. E., & Hall, S. E. (2007). Adaptation of central pitch-specific mechanisms. *Perception*, 36: 918–930.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27: 283–296.

Qiu, L., & Benbasat, I. (2005). Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and Three-Dimensional Avatars. *International Journal of Human-Computer Interaction*, 19(1): 75–94.

Reips, U. D., & Funke, F. (2008). Interval level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavioural Research Methods*, 40(3): 699-704.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology*, 14: 2119–2123.

Riding, D., Lonsdale, D., & Brown, B. (2006). The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women. *Journal of Nonverbal Behavior, 30*: 55–61.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion,* 9(2): 260-264.

Saxton, T., Caryl, P., & Roberts, S. C. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: the ontogeny of mate choice. *Ethology,* 112: 1179–1185.

Scherer, K. R. (1972). Judging Personality from Voice - Cross-Cultural Approach to an Old Issue in Interpersonal Perception. *Journal of Personality,* 40(2): 191-210.

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience,* 12: 508–514.

Schotz, S. (2007). *Acoustic analysis of adult speaker age. Speaker Classification I:* Berlin, Heidelberg: Springer.

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., . . . & Zaske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, 18: 684–688.

Siann, G. (1994). *Gender, sex and sexuality: Contemporary psychological perspectives.* London: Taylor & Francis.

Skorowski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin,* 105(1): 131-142.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J. M., Burt, D., & Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition, [Research Support, Non-U.S. Gov't],* 127(1): 105-18.

Thompson, J. D. (1967). *Organizations in action: social science bases of administrative theory*. New York: McGraw-Hill.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences,* 12(12): 455-60.

Tsankova, E., Aubrey, A. J., Krumhuber, E., Möllering, G., Kappas, A., Marshall, D., & Rosin, P.L. (2013), Facial and Vocal Cues in Perceptions of Trustworthiness. *Computer Vision - ACCV 2012 Workshops Lecture Notes in Computer Science*, 7729: 308-319.

Tuomi, S. K., & Fisher, J. E. (1979). Characteristics of simulated sexy voice. *Folia Phoniatrica et Logopaedica*, 31: 242–249.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A,* 43: 161–204.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, 21(6): 779-85.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal Orientation and the Accuracy of Personality Judgments. *Journal of Personality*, 71: 267–295.

Vukovic, J., Jones, B. C., DeBruine, L. M., Feinberg, D. R., Smith, F. G., Little, A. C., . . . & Main, J. (2010). Women's own voice pitch predicts their preferences for masculinity in men's voices. *Behavioral Ecology,* 21(4): 767-772.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428: 557–561.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366: 1702–1725.

Welling, L. L. M., Jones, B. C., DeBruine, L. L. M., Smith, F. G., Feinberg, D. R., Little, A. C., & Al-Dujaili, E. A. S. (2008). Men report stronger attraction to femininity in women's faces when their testosterone levels are high. *Hormones and Behavior*, 54: 703-708.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psycholical Science* 17: 592–598.

Winquist, L. A., Mohr, C. D., Kenny, D. A. (1998). The female positivity effect in the perception of others. *Journal of Research in Personality*, 32: 370–388.

World Medical Association (initiated 1964, amended 1975, 1983, 1989, 1996, 2000, 2002). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. In: Griffin, J., P., & O'Grady, J. (Eds.) (2006*). The Textbook of Pharmaceutical Medicine* (Appendix 1), Oxford: Blackwell Publishing.

Wyer Jr., R. S., & Watson, S. F. (1969). Context effects in impression formation. *Journal of Personality and Social Psychology,* 12(1): 22-33.

Young, A.W., & Bruce V. (2011). Understanding person perception. *British journal of psychology*, 102: 959-74.

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences,* 17(6): 263 – 271.

Zaidel, D. W., Bava, S., & Reis, V. A. (2003). Relationship between facial asymmetry and judging trustworthiness in faces. *Laterality*, 8(3): 225-232.

Zebrowitz, L. A. (1996). Physical appearance as a basis for stereotyping. In: Macrae C.N., Hewstone M., Stangor C., editors. *Foundation of stereotypes and stereotyping, pp. 79-120.*. New York: Guilford Press.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul*? Boulder, Colorado: Westview Press.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc,* 1(3): 204-23.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origins. *Personality and Social Psychology Bulletin,* 28: 238–249.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *Journal of*

Personality and Social Psychology, [Research Support, N.I.H., Extramural], 98(2): 175-89.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, 2(3): 1497-1517.

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to “Bad Genes” and the Anomalous Face Overgeneralization Effect: Cue Validity, Cue Utilization, and Accuracy in Judging Intelligence and Health. *Journal of Nonverbal Behavior*, 28: 167–185.

Zebrowitz-McArthur, L. A., & Montepare, J. M. (1989). Contributions of a babyface and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior*, 13(3): 189-203.

Zuckerman, M., & Driver, R. E. (1989). What Sounds Beautiful Is Good - the Vocal Attractiveness Stereotype. *Journal of Nonverbal Behavior,* 13(2): 67-82.

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior.* 17(2): 119–135.