Trust in the Voice is Highly Context Specific

Isar Syed

School of Humanities and Digital Sciences, Tilburg University

840900-B-18: Bachelor thesis Liberal Arts and Sciences

Prof. dr. Jean Vroomen

13 December 2021

Abstract

Research into the nature of trustworthiness judgments has previously focused on the different aspects. Parameters such as vocal pitch and facial Width-to-Height Ratio (fWHR) have been identified to exert an influence on perceived trustworthiness (Barclay, 2017; Geniole, 2014), but studies on these parameters more often than not only focus on one of these factors; either the face or the voice. Yet, these factors tend to coincide in daily life and are seldom separated. It could therefore be expected that the face and the voice influence one another in the assimilation of trust judgments. Additionally, trust may differ between situational contexts. This study uses a multimodal design to investigate the link between the fWHR and vocal pitch across neutral and mating-related contexts in trust and attraction. Results show that perceived trustworthiness is highly dependent on situational context, as lower-pitched male voices were perceived as the least trustworthy, but also as the most attractive in mating-related circumstances. The fWHR, vocal pitch, and sentence content all influence trust and attraction in different ways, as well as interact with one another.

A fundamental part of human relationships and interaction is our capacity to trust and distrust others. Trust has previously been defined as the feeling or idea that others will behave fairly towards you (Ward, 2016), and humans have adaptively evolved to make estimations on whom to trust and whom to distrust (Zebrowitz, 2017). Upon the first encounter, there are many factors that contribute to our perception of one's trustworthiness and other personality traits, two of which are the voice and the face of the persons whom we encounter. A large body of research is devoted to investigating these facets separately (Alaei, 2020; Feinberg, 2005; Stirrat, 2010). For the voice, parameters such as higher pitch have been identified to increase perceived dating-related trustworthiness (Belin, 2017), while another study finds that

higher pitch decreases trustworthiness (Schirmer, 2020), and these findings tend to differ between men and women (Barclay, 2017). Barclay & O'Connor (2017) report that listeners indicated higher general trustworthiness of high-pitched male voices and low-pitched female voices.

New insights into these two facets of human trust interaction are of importance to our understanding of human interpersonal relationships. In the research literature, several sources provide information about impressions derived from the voices of strangers. Previous studies have tested multiple parameters of the voice to understand their effect on the perceived trustworthiness and attractiveness, such as fundamental frequency (F0) (Feinberg, 2005), or accent (Jiang, 2020). Literature on the characteristics of a trustworthy or attractive voice, however, reports inconsistent results. The most important findings include an influence of the average pitch of the voice (Schild, 2019). Male speakers with a lower F0 were judged as being more trustworthy than male speakers with a higher F0 in economic settings, yet no such effect was found for female speakers. Interestingly, it might be the case that females judge both male and female voices as more attractive when the pitch was raised ((F0+20Hz), but this effect was found in reverse in another study (Suire, 2019). Furthermore, it has been suggested that men with relatively more masculine (i.e. lower) voices are judged to be more likely to commit adultery (Suire, 2019), a result confirmed in other studies (Schild, 2019), and that women who possess higher pitched voices are judged to be less likely to romantically cheat (Barclay, 2017). These findings hint that vocal pitch may be related to trustworthiness in mating-related contexts.

Similar attention has been given to the search for facial features that are indicative of social trait judgments. Most searches for correlations between facial features and social traits have resulted in insignificance, but there is a relatively large body of literature focusing on a particular ratio between the eyes and the mouth, and face width the so-called facial Width-to-

Height Ratio (fWHR) (Geniole, 2014; Geniole, 2015; Haselhuhn, 2017;). People are consistently found to be judged as more dominant when their fWHR is digitally manipulated to be larger (i.e., smaller distances between the eyes and mouth) (Stirrat, 2010). They further add that men with greater facial width were more likely to exploit the trust of others in trust games. This ratio is thus believed to influence interpersonal trust dynamics.

Further, it has been demonstrated that the facial Width-to-Height ratio influences trustworthiness differently across different circumstances. As explained before, the fWHR ratio is a ratio the distance between the eyes and the mouth, and face width (Geniole, 2015). Generally, people with smaller distances between the eyes and top lip, and broader distances between the eyes are judged to be more dominant at first glance (Geniole, 2014). Research further indicates that fWHR influences the perception of dominance and the perception of emotional intensity (Merlhiot, 2021). Importantly, fWHR only exhibited effect on dominance in dominance-based emotional facial expressions such as anger. This indicates that people with a larger fWHR are likely to be perceived as more dominant in situations that are sensitive to dominance, such as an argument. Men with smaller fWHRs are, however, reported to be perceived as having greater integrity, which suggests that people may be chosen for different roles (i.e., leadership or romantic roles) based on superficial characteristics (Haselhuhn, 2017).

In everyday conversation, however, the face and voice usually coincide, suggesting a feature of multimodality to social trait judgments. This multimodality has been identified by Tsankova et al. (2015), who subjected participants to video of people saying ''Hello, my name is Jo''. Their design included both consistent and inconsistent combinations of facial-vocal communication (i.e., a trustworthy voice and a trustworthy face, or an untrustworthy voice and a trustworthy face). Participants were then asked to rate how trustworthy the speaker is. They report that in the assimilation of trustworthiness judgments, the face has a

superiority effect over the voice. These findings suggest that trustworthiness impressions are more influenced by the face than the voice. This is the only study on multimodal trust impressions that was found.

Still, daily conversations tend to be more variant in subject matter and complexity than is often assumed in research: as is shown, little is known about how the interaction between facial and vocal cues might be influenced by contexts, for instance neutral and mating-related contexts suggested by Barclay et al. (2017) and Schild et al. (2019). Given that so many influences and effects are at play, it is clear that trust impressions from daily conversation are more complex than is usually assumed in research. The current study bridges literature between facial cues in the form of the fWHR with literature on vocal cues in the form of F0 and sentence content and investigate their relationship, as well as investigate the link between trustworthiness and attractiveness. It will do so by conducting unimodal and multimodal experiments measuring trustworthiness judgments from models with varying fWHRs and pitch manipulations, as well as differentiating between neutral and mating-related contexts (Study 1). It is expected that differences in fWHR will manifest in differences in trustworthiness across contexts, as in neutral contexts, larger ratios, being associated with higher dominance (Geniole, 2014), are expected to be found to be more trustworthy. In mating-related contexts, smaller ratios are anticipated to be found to be more trustworthy, as larger ratios are associated with greater integrity (Haselhuhn, 2017). With regard to the voice, it is hypothesized that lower voices are perceived as more trustworthy in neutral but not mating-related contexts, as suggested by Barclay et al. (2017) and Schild et al. (2019). Of most interest to the current study, it is anticipated to find both the fWHR and vocal pitch to be interacting with the context of the sentence. larger fWHRs, being associated with dominance (Geniole, 2014), are expected to be perceived as more trustworthy in neutral contexts, and the same is expected of lower-pitched voices. However, it should not be

surprising to find these effects differing for men and women, as well as different contexts, as it is expected that male speakers will be evaluated as less trustworthy in dating, but not neutral contexts, than female speakers.

Study 2, described further, was designed to measure attractiveness of the same stimuli.

Study 1: Trustworthiness of the face and voice

The aim of this experiment was to investigate the effects of vocal pitch and content, and facial Width-to-Height ratio on trustworthiness of male and female speakers. Stimuli were created by recording and photographing four models and these stimuli were presented to participants both unimodally and audiovisually.

Methods

Procedure

This study was conducted at the Cognitive Neuroscience laboratory at Tilburg University. The participants were recruited through the participant system of Tilburg University. A total of 33 (7 male, 26 female) participants with a mean age of 19.8 years old (SD=2.073) were invited to the lab, where they, after providing written consent, completed the first part which entailed testing the voices and faces on trustworthiness unimodally. The experiments were conducted in sound-proof cabin on a [MONITOR NAME] monitor with a refresh rate of [REFRESH RATE] Hz. The audio was provided through Sennheiser HD201 headphones at an average volume of 45 dB. The voice clips were presented first, and after hearing the clip, participants were asked ''How trustworthy did you experience this voice?''. They then could communicate their judgment by pressing one of the number keys on the

keyboard (1 to 9). Upon finishing these trials, the faces were shown, and participants were asked how trustworthy they experienced the face, with the same instructions as with the voices. After a short break, the audiovisual part of the experiment was presented. The faces were now shown simultaneously to the voices, and participants were asked again how trustworthy they experienced the voice. For both the unimodal and audiovisual parts, stimuli were presented per model in a blocked fashion; first all stimuli of model 1 were presented, followed by all stimuli of model 2, and so on. It usually took participants about 30 to 40 minutes to complete these two parts. Lastly, after another short break, participants completed Study 2 (described further). The full duration of all three parts was circa 50 minutes. Upon completion, the participants were thanked and compensated with 1 participation credit.
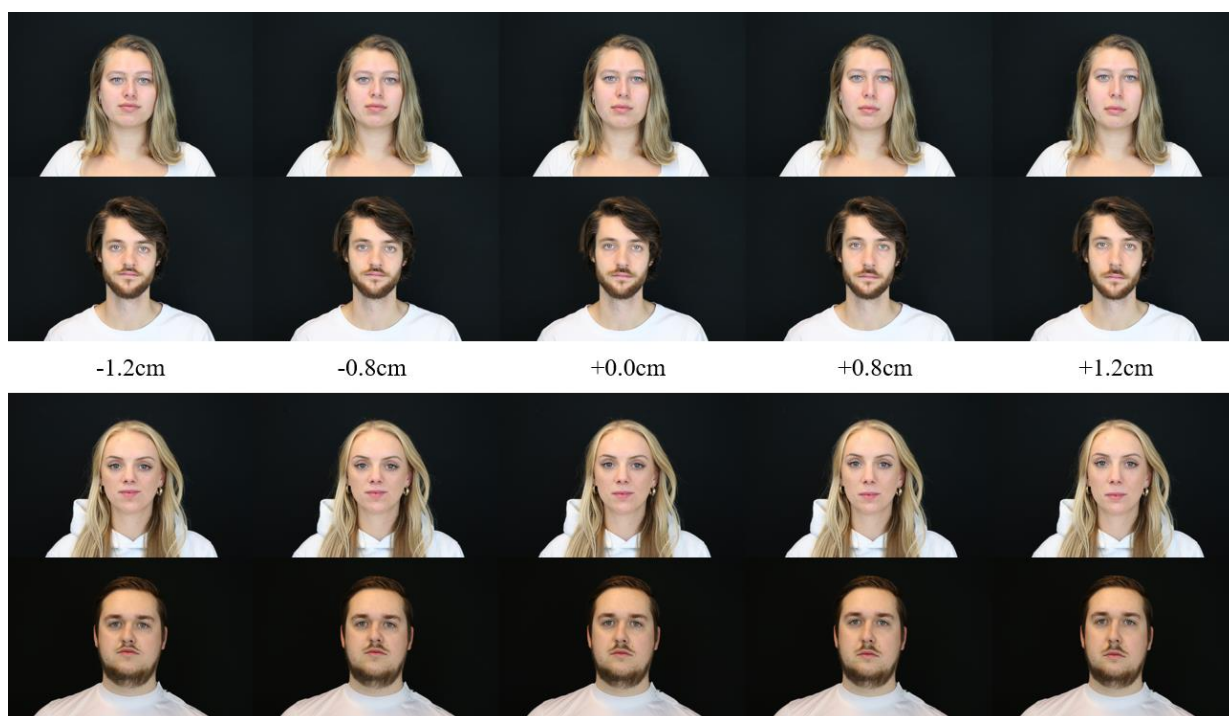
Stimuli

For the facial stimuli, pictures were taken of two men and two women. All models were Caucasian Dutch speaking University studen. Photographs were taken from a set range of about 100cm using a [NAME OF CAMERA] in a well-lit room with a black background. The models were instructed to refrain from displaying any emotional facial expressions. Next, the photographs were imported into Photoshop (Adobe Inc, 2020), where manipulations to the fWHR were applied. To remain consistent, the fWHR was manipulated by changing the distance between the eyes and mouth in steps of 0.4cm in each direction. However, both the -0.4cm and +0.4cm manipulations were discarded due to these manipulations being too similar to the originals, and because of time constraints in the experiment. Thus, the manipulations consisted of -1.2cm, -0.8cm, +0.0cm, +0.8cm, and +1.2cm alterations. These distances are based on the distance calculations by Photoshop and were chosen due to larger steps being intuitively perceived as unnatural. Because the pictures

were taken from a constant distance, complying with the distance measures provided by Photoshop did not result in inconsistent manipulations across models. In addition to manipulations in the fWHR, changes were made to the forehead, chin height, nose height, and face width when necessary to accommodate the changes in the distance between the eyes and mouth. All visual stimuli used in this study can be found below (Fig. 1). In total, there were 4 models each manipulated with 5 different fWHRs, resulting in 20 photographs.

Figure 1

fWHR Manipulations on all models

Auditory stimuli were recorded from the same models to ensure natural audio-visual congruency. Each model was instructed to speak four sentences in a neutral tone of voice. Sentences were recorded using a [NAME OF MICROPHONE]. Two of the sentences were neutral in content ("De bus naar de buitenwijk vertrekt over vijftien minuten", and "Zijn vriendin kwam uit het vliegtuig"). The other two sentences were mating-related ("Wil je

volgende week wat met mij drinken?", and "Wat heb je mooie ogen"). An overview of the full sentences, their English translation, and situational context is provided in Table 1.

Table 1

Sentences with their closest English translation, and substantive context

Sentence	Full sentence	English translation	Context
I	De bus naar de buitenwijk vertrekt over vijftien minuten.	The bus to the outskirts leaves in fifteen minutes.	Neutral
II	Zijn vriendin kwam uit het vliegtuig.	His girlfriend came out of the airplane	Neutral
III	Wil je volgende week wat met mij drinken?	Would you like to have a drink with me next week?	Mating-related
IV	Wat heb jij mooie ogen.	You have beautiful eyes.	Mating-related

The recordings were manipulated using the voice editing software Praat (Boersma, 2013). For each recording, the average F0 (corresponding to pitch) was increased and decreased by 20Hz, based on previous research (Suire, 2019). There was a total of 4 models speaking 4 sentences that were manipulated in 2 ways, resulting in 8 audio clips per model, so 32 clips in total. An overview of the mean F0 of each model's voice for each sentence can be found in Table 2.

Table 2

Mean F0 for each model and each sentence in Hertz (Hz)

Sentence	Pitch	GN (Male)	MM (Male)	DE (Female)	LH (Female)
I	Low	93.34 Hz	87.23 Hz	167.77 Hz	140.86 Hz
	High	122.23 Hz	121.26 Hz	208.27 Hz	180.56 Hz
II	Low	99.42 Hz	95.26 Hz	200.63 Hz	133.67 Hz
	High	123.77 Hz	133.33 Hz	239.62 Hz	172.67 Hz
III	Low	100.09 Hz	93.52 Hz	159.52 Hz	144.76 Hz
	High	135.51 Hz	130.48 Hz	204.38 Hz	183.89 Hz
IV	Low	95.04 Hz	87.10 Hz	181.60 Hz	149.28 Hz
	High	144.29 Hz	126.23 Hz	222.44 Hz	188.49 Hz

An experiment was created in the psychology experiment software Opensesame (Mathôt, 2012) that presented the manipulated audio files and photographs separately, upon which participants could rate the trustworthiness of each stimulus on a Likert scale ranging from 1 to 9 (1 = Not trustworthy at all, 9 = highly trustworthy). Data collected from these trials would provide insight into trustworthiness score from the stimuli unimodally. Next, participants were shown the same stimuli audiovisually. This part would test the interaction between the audio and visual stimuli. Similar to the first and experiment, this experiment consisted of trials in which stimuli were presented and trustworthiness ratings were given by the participants. In each trial, participants were asked to rate the trustworthiness of the voice on a 9-point Likert scale (1 = Not trustworthy at all, 9 = highly trustworthy), while paying attention to the face. This resulted in a 4 (models) x5 (facial manipulations) x8 (vocal stimuli) design.
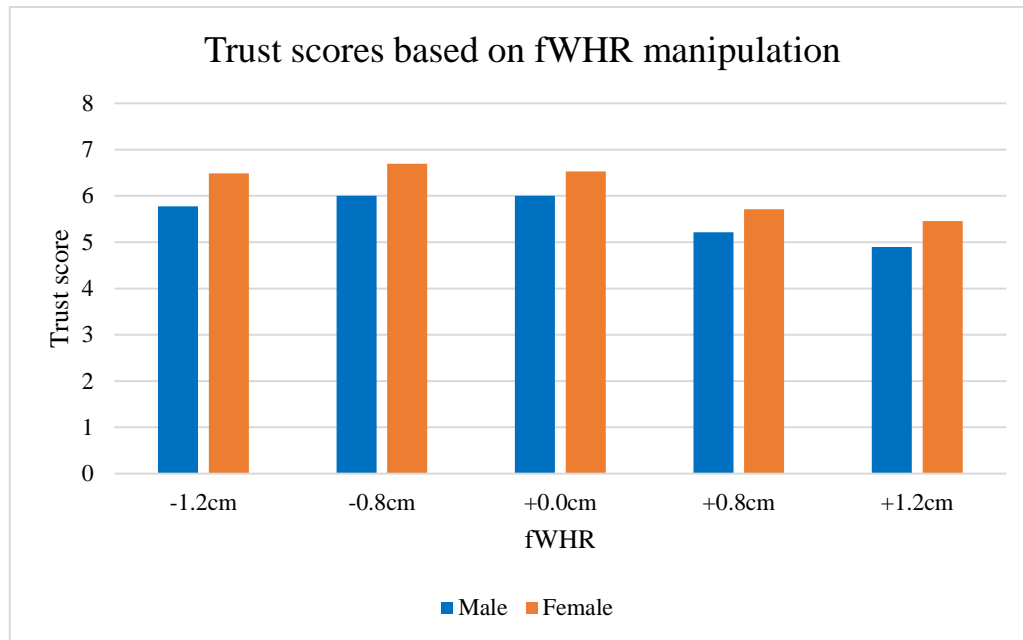
Data analysis

The unimodal trust scores of the voice were analysed in a 2 (Content: Neutral or Mating-related)*2 (Pitch: High (F0+20Hz) or low (F0-20Hz)*2 (Gender of the model: Male or Female) MANOVA. For the unimodal visual stimuli, a 2 (Gender of the model: Male or Female) * 5 (fWHR manipulations) MANOVA was conducted. Audiovisual data were analysed using a 2 (Content: Neutral and Mating-related) * 5 (fWHR manipulations) * 2 (Pitch: Low (F0-20Hz) and High (F0+20Hz)) * 2 (Gender of the speaker: Male and Female) MANOVA. Significant results from the multivariate tests were further examined using paired t-tests.

Results

Figure 2, Figure 3, and Figure 4 below visualize the unimodal data of this study. Analysis of the unimodal visual experiment measuring trustworthiness of the face revealed statistically significant effects for the fWHR manipulations for both men ($F(4,160)=5.166$, $p=0.001$, $\eta 2=0.114$) and women ($F(4,160)=5.720$, $p<0.001$, $\eta 2=0.125$). For both sexes, smaller ratios were perceived as less trustworthy, as shown by Tukey contrasts (Males: M(fWHR +1.2cm)= 4.894 vs. M(fWHR -1.2cm)=5.773, Females: M(fWHR +1.2cm)=5.445 vs. M(fWHR -1.2cm)=6.485). A complete overview of trustworthiness ratings per fWHR can be found in Figure 2.

Figure 2

Trust scores based on fWHR manipulation.

With regard to the vocal stimuli, data analysis revealed significant effects for Content ($F(1,32)$=6.803, p=0.014, η2=0.175), as mating related sentences were seen as less trustworthy than neutral contexts (M(Mating)=6.0246, SD= 0.955 vs. M(Neutral)=6.3283, SD= 1.006), and for Pitch ($F(1,32)$=5.411, p=0.026, η2=0.145), as low-pitched voices were perceived as more trustworthy than high-pitched voices (M(Low)=6.0227, SD= 1.077 vs. M(High)=6.3302, SD= 0.911) (Fig. 3 and Fig. 4). This finding indicates that there are differences in trust based on the content of the sentence and on vocal pitch. The interaction Content*Pitch was also found to be significant ($F(1,32)$=13.933, p=0.001, η2=0.303), as low-pitched mating-related sentences were found less trustworthy in comparison to low-pitched mating-related sentences (M(Low-Mating)=5.7235, SD= 1.156 vs. M(High-Mating)= 6.3220, SD= 1.136). Importantly, no such difference was detected for high pitch. This highlights that the effect of the content of the sentence is dependent on vocal pitch. Further, the interaction

effect Content*Pitch was possibly still mediated by Gender, but the two-way interaction effect Content*Pitch*Gender was found to only be marginally significant (F(1,32)=3.852, p=0.058, η2=0.107). This two-way interaction suggests that trustworthiness of a voice is based on its pitch and different for certain contexts, and that this difference is also dependent on the gender (male or female) of the speaker.

Figure 3

Trust scores for high and low pitched sentences in neutral contexts
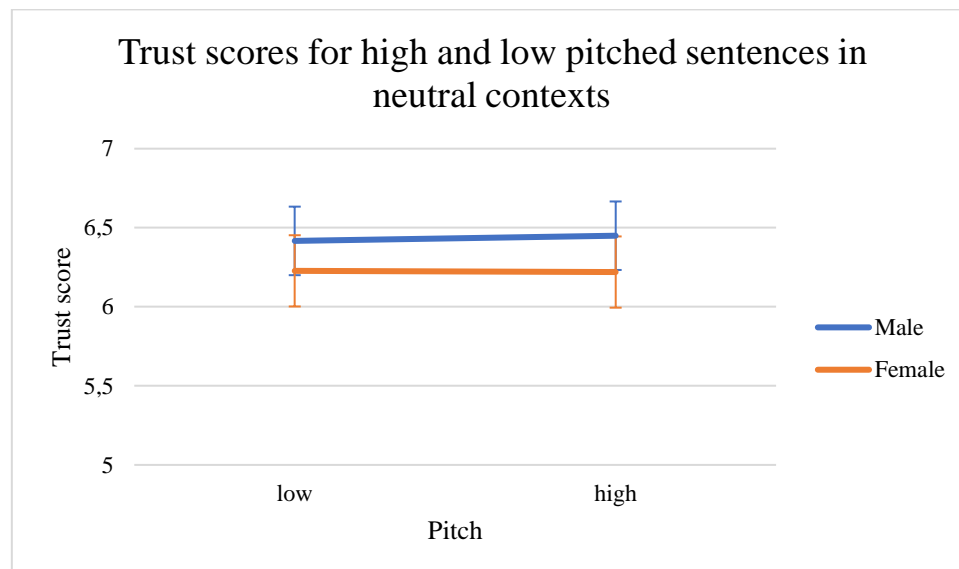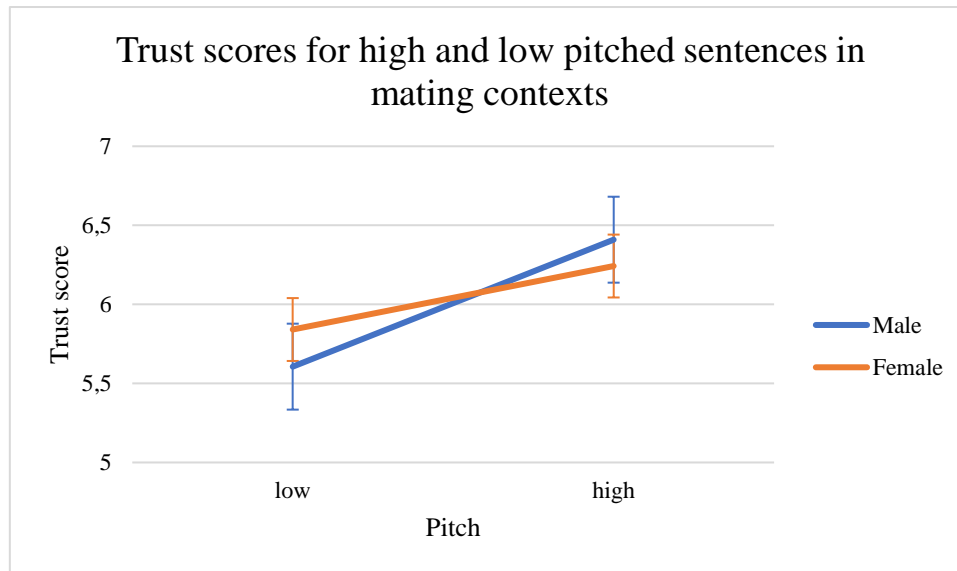
Figure 4

Trust scores for high and low pitched sentences in mating-related contexts

Analysis of the audiovisual data produced the graphs shown below (Fig. 5 and Fig. 6). A clear difference can be seen between trust scores and the content of the sentence. The datapoints are much more dispersed in the mating-related graph than in the neutral graph, showing that differences in trustworthiness are revealed by the content of the sentence ($F(1,32)=15.383$, $p<0.001$, $\eta2=0.325$). Pitch was also found to be significant ($F(1,32)=6.483$, $p=0.016$, $\eta2=0.168$), where lower pitched voices were found to be less trustworthy than high-pitched voices (M(low)$=5.9553$, SD$= 1.014$ vs. M(High)$=6.2015$, SD$= 0.920$) across contexts. The fWHR was also significant on trustworthiness ($F(4,29)=6.047$, $p=0.001$), $\eta2=0.455$) . The graph shows higher trust scores for larger than smaller fWHRs (M(fWHR-1.2cm)$=6.2131$, SD$= 0.952$ vs. M(fWHR +1.2cm)$=5.7320$, SD$= 1.154$). This effect was similar for both genders, pitches, and contexts.

A number of interaction effects were also discovered. As can be seen in comparison between Fig. 5 and Fig. 6, differences in trustworthiness based on pitch become more

noticeable when differences in content are introduced ($F(1,32)=4.287$, $p=0.047$, η2=0.118): lower pitched voices were perceived as less trustworthy in mating-related contexts. Further, a strong relationship was discovered between gender and content ($F(1,32)=14.684$, $p=0.001$, η2=0.315). As shown in the graphs, women are perceived as more trustworthy in mating-related, but not neutral contexts, than men. Lastly, the graphs hint that trustworthiness of the person depends on the context and differences in fWHR, and that these differences might also depend on the gender of the person. The two-way interaction Content*Face*Gender was also found to be significant ($F(4,29)=3.140$, $p=0.029$, η2=0.302), as the distance between datapoints for each fWHR is greater for male speakers than for female speakers, and this difference seems to be more pronounced in mating-related contexts.

Figure 5

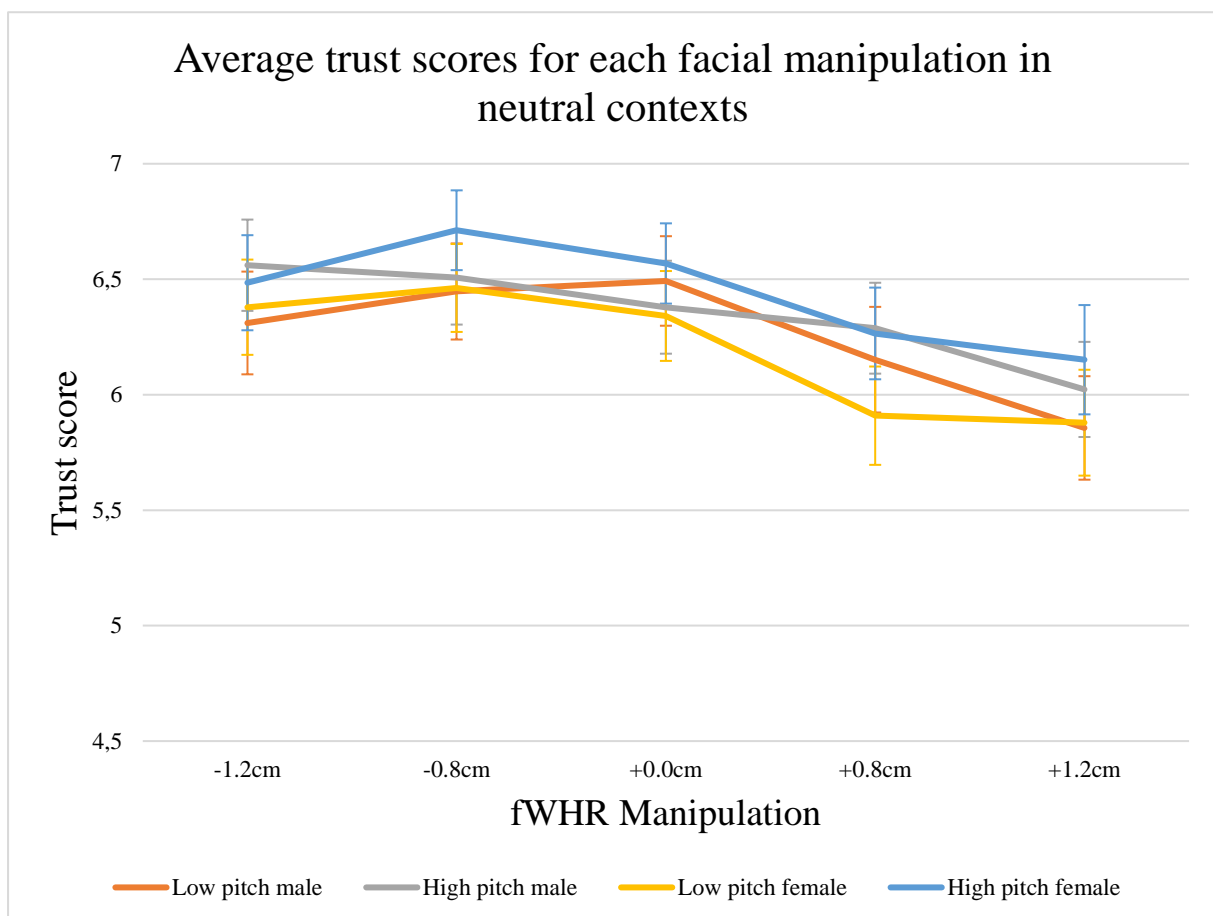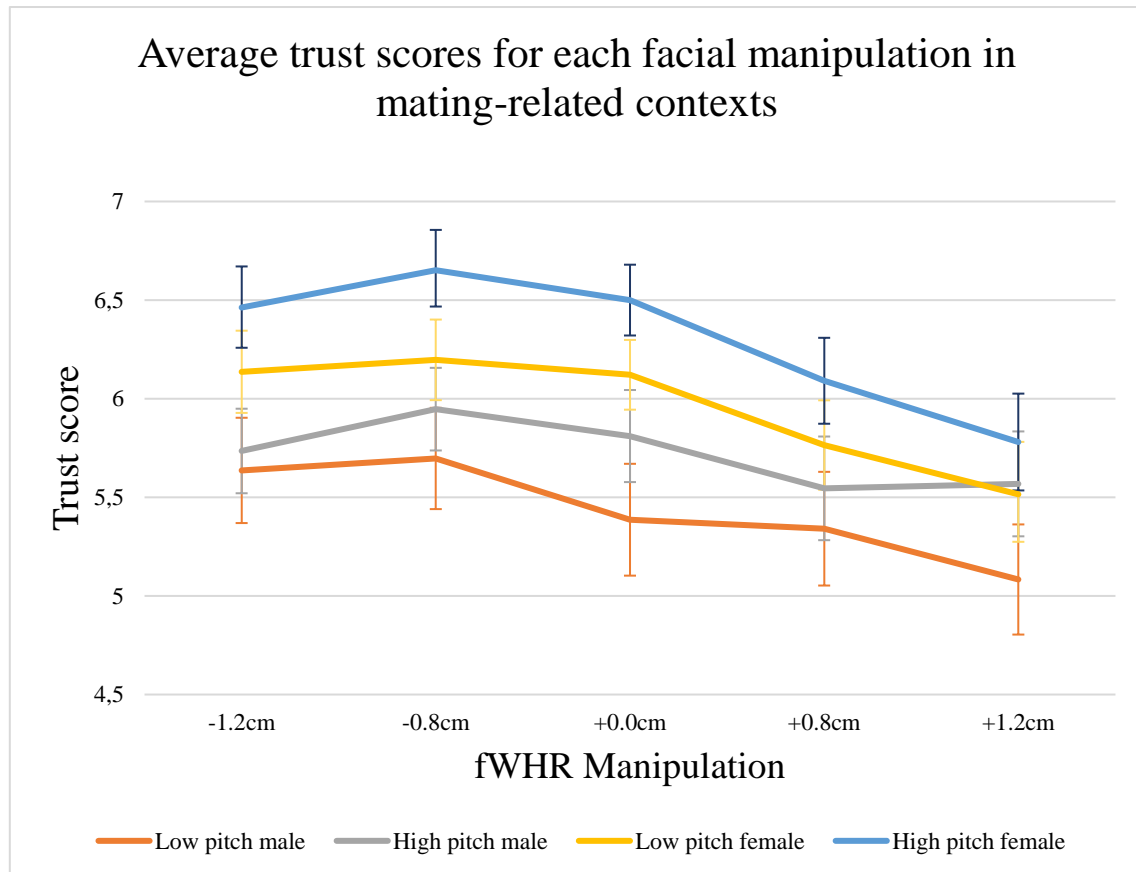*Average trust scores for each fWHR manipulation in neutral contexts*

Figure 6

Average trust scores for each fWHR manipulation in mating-related contexts

Average trust scores for each facial manipulation in mating-related contexts

Discussion

    The purpose of the current experiment was to bridge findings between facial cues in the form of the fWHR with findings on vocal cues in the form of F0 and sentence content and investigate their interaction. The results of this study demonstrate that trustworthiness of a person is strongly influenced by context, vocal pitch, face, and gender of the speaker, and that these factors interact. Most importantly, low pitched voices of males in mating-related context were perceived as being most untrustworthy. These results are very much in line with the suggestions by Schild et al. (2019) and Barclay et al. (2017), that trust is mediated by vocal pitch as well as the context of the interaction. Further, the results show that the fWHR

exhibits an influence on trust. It is not surprising to find that the fWHR has an effect, noting the supremacy effect of the face over the voice in trust judgments identified by Tsankova et al (2015).

In accordance with the hypothesis, the fWHR displayed an effect on perceived trustworthiness. Larger ratios (i.e. smaller distances between the eyes and mouth) were related to higher trust scores in neutral contexts. However, smaller ratios (i.e. greater distances between the eyes and mouth) were not found to increase trustworthiness in mating-related contexts. It was, in fact, larger ratio's that were perceived as more trustworthy in this context.

Study 1 demonstrates that trustworthiness in the voice is affected by pitch, gender and content: male low-pitched voices in mating-related context were judged to be least trustworthy. Is this a specific effect on trustworthiness of the voice, or does this generalize to other attributes of the voice, such as its attractiveness? Can an untrustworthy voice be attractive? In Study 2, we examined this intriguing question by asking participants to rate the same stimuli on their attractiveness.

Study 2: Unimodal Attractiveness

The aim of this experiment was to investigate the effects of vocal pitch, content, and fWHR on attractiveness of male and female speakers. Attractiveness was investigated to see whether or not differences in trustworthiness could be explained through attractiveness.

We expected that larger fWHRs would be perceived as more attractive for males (Kramer, 2017). Also, lower-pitched voices will be seen as more attractive for male speakers, and

female speakers were expected to be regarded as more attractive with a higher vocal pitch, based on trust (Hirschberg, 2018).

Methods

This experiment was identical in structure to the unimodal part of Study 1. The same stimuli and instructions were given to participants. The difference between the unimodal part of the first study and the current experiment is that instead of trustworthiness, this experiment tested attractiveness of the voice and face. Again, first the voices were presented and participants were asked ''How attractive did you experience this voice?'', and then the faces were shown with again the question ''How attractive did you experience this face?''. The same instructions as before applied here. This experiment was conducted in addition to the first study on the last 18 participants of the subject pool. These participants (16 female, 2 male) had an average age of 19.56 (SD=2.036).

Results

Group-averaged scores on attractiveness of the face are shown in Figure 7. Data were analysed as before. Analysis of the data revealed significant effects for the fWHR for both men ($F(4,75)=5.049$, $p=0.001$, $\eta2=0.212$) and women ($F(4,75)=10.547$, $p<0.001$, $\eta2=0.360$) on attractiveness. Tukey contrasts showed that larger ratios were generally perceived as being more attractive across genders (M(fWHR +1.2cm)=3,5625 vs. M(fWHR -0.8cm)=6.6095).
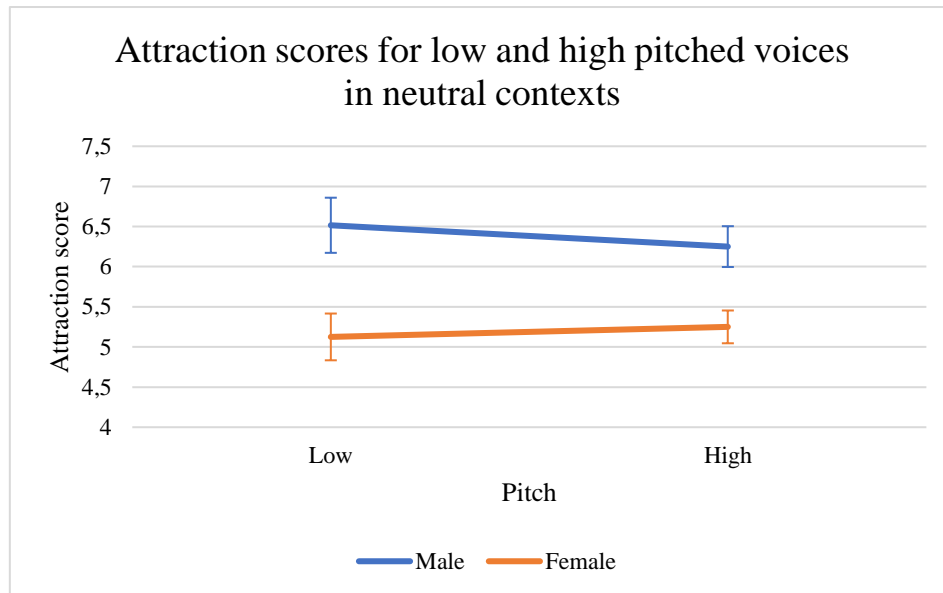
Figure 7

Attraction scores based on fWHR.
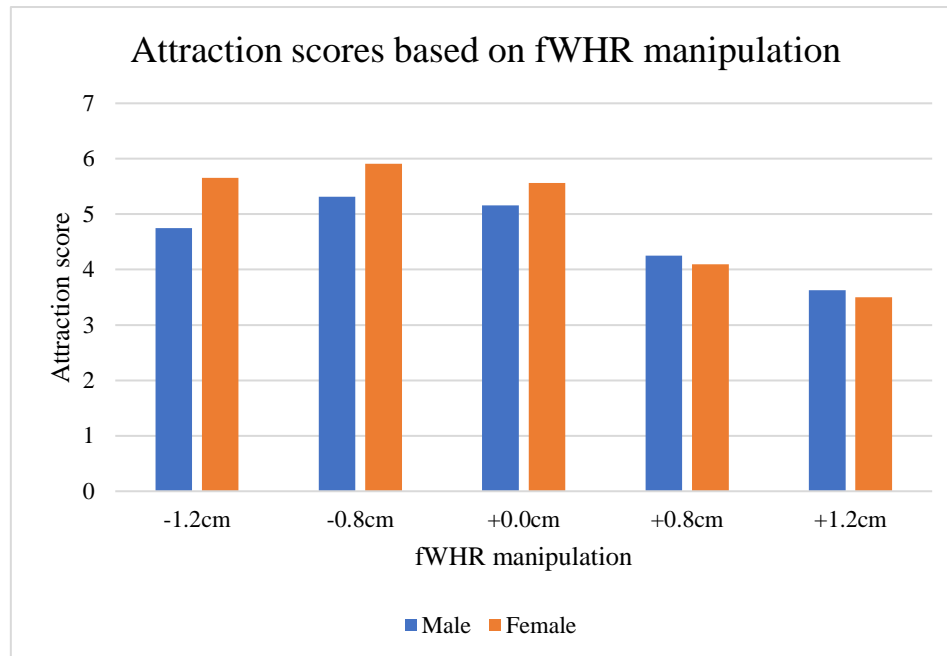
A 2(Gender) x 2 (Pitch) MANOVA showed a significant effect of the gender of the speaker (F(1,15)=16.791, p=0.001, η2=0.528), as male voices were judged as more attractive than the female voices (M(Male)=6.3711, SD= 1.023 vs. M(Female)=5.1836, SD= 0.759) (Fig. 8 and Fig. 9). No further main effects were found for vocal pitch or sentence context. However, there might be an interaction effect between vocal pitch and gender, as low-pitched voices are judged as more attractive for male speakers than low-pitched voices for female speakers (M(Male)=6.5547, SD= 1.340 vs. M(Female)=5.1094, SD= 1.028), but this interaction did not reach statistical significance (F(1,15)=3.796, p=0.070, η2=0.202).

Figure 8

Attraction scores for low and high pitched voices in neutral contexts

Attraction scores for low and high pitched voices in neutral contexts

Figure 9

Attraction scores for low and high pitched voices in mating-related contexts

Attraction scores for low and high pitch voices in mating-related contexts
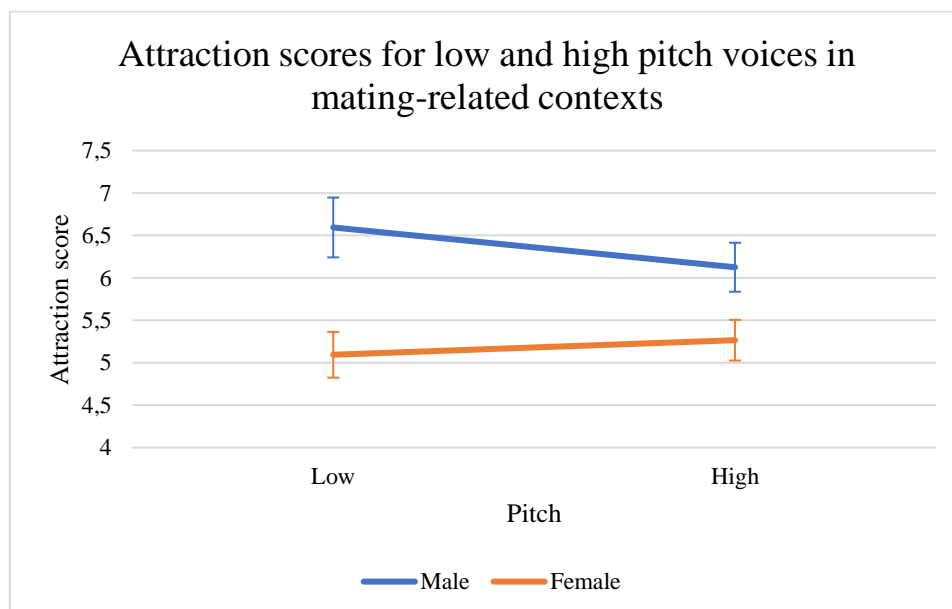
In comparison of unimodal trust and attraction data, few differences could be found in the fWHR; for both trust and attractiveness, larger fWHRs tend to be viewed as more

favourable for both males and females. The voice showed more interesting differences, as a repeated measures ANOVA on a combination of trust and attraction scores revealed that there is a significant two-way interaction effect between the experiment (trust or attractiveness), gender, and vocal pitch in mating-related contexts ($F(1,15)=11.716$, $p=0.004$, $\eta2=0.439$). Male low-pitched voices were seen as more attractive ($M=6.5938$, SD= 1.411) and less trustworthy ($M= 5.6061$, 1.357) on the same scale. This effect was inverted for high pitch, as male high-pitched voices were simultaneously seen as less attractive ($M= 6.1250$, SD= 1.155) and more trustworthy ($M= 6.1875$, 0.981)), however this difference for high pitch was not significant ($t(15)=0.161$, $p= 0.874$). Comparison of the graphs for trust and attractiveness score of the voice in mating-related contexts suggest the score given to the voice is thus dependent on vocal pitch, and that this differs whether attractiveness or trustworthiness is being measured, and that these scores differ between men and women. This effect is visualized below (Figure 10).

Figure 10

Trust and attractiveness scores for low and high pitched sentences in mating-related contexts
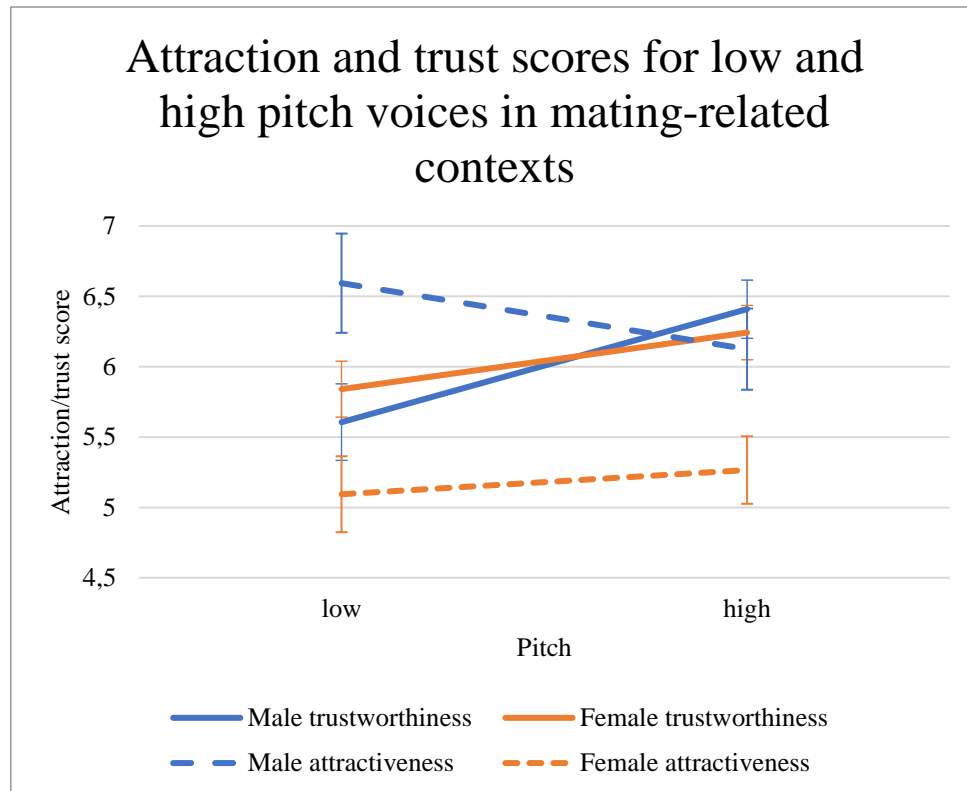
Discussion

This experiment was conducted to investigate the effects of the content of a sentence, vocal pitch, and gender of a speaker, on perceived attractiveness. The results show a strong relationship between the fWHR and attractiveness. For both men and women, larger ratios (i.e. smaller distances between the eyes and mouth) are judged as more attractive. For males, this might be because an identified factor in facial attractiveness is sexual dimorphism (Little, 2011), and the fWHR is a sexually dimorphic, testosterone-linked trait (Stirrat, 2010). The fWHR might therefore have been a cue of sexual maturity. However, this claim has been disputed (Kramer, 2017). For females, no explanation could be found in the existing literature. This might therefore be an interesting topic for future research.

Male voices were judged to be more attractive than female voices, with little effect displayed by vocal pitch. It might, however, be the case that vocal pitch is involved in an interaction with gender. This finding would suggest that attractiveness of the vocal pitch is dependent on the gender of the speaker. This means that low-pitched male voices might be perceived as more attractive than high-pitched male voices, while this could be different for female voices, and it is therefore invalid to claim that either higher or lower-pitched voices are more attractive irregardless of gender. As this finding is close to significance, future research might reveal more of this suspicion.

Of great importance to the study, results from the comparison between trust and attractiveness suggest that female voices tend to be more trustworthy than attractive, and higher-pitched voices score higher for both measurements (Fig. 10). More interestingly, low-pitched male voices are more attractive but less trustworthy in mating-related contexts. High-pitched male voices are simultaneously more trustworthy and less attractive. This demonstrates that trust and attractiveness do not necessarily coincide.

In accordance with the hypothesis, larger fWHRs were found to be more attractive for both genders. Further, lower pitched male voices were indeed found to be more attractive, and female voices were also regarded as more attractive in higher vocal pitch.

General Discussion

The goal of these studies was to investigate the nature of trustworthiness through multimodal investigation. This research focused on the facial Width-To-Height ratio (fWHR) and vocal pitch as main influences on trust, but also incorporated different contexts to measure the effect on the content of the message. Additionally, these same variables were used to measure their influence on attractiveness unimodally. Results of these studies suggest

that trust in the voice is highly context specific: people tend to be more careful in mating-related contexts than in neutral contexts. This effect seems to be very sensitive to the context, as a simple change in sentence exerted a measurable influence. More so than vocal pitch, gender, or fWHR variation, context seems to be an effective predictor of trust in human interaction. Note, however, that trust could not be explained through attractiveness. For the voice, only the gender of the speaker was found to have a significant effect, and vocal pitch seemed to have no influence. This was, of course, to be expected, seeing as the participant pool consisted of mostly women. For the face, the fWHR was found to be a significant factor in both trust and attractiveness judgments, yet, it is difficult to say what the relation between the two is as they might be separate psychological mechanisms.

The interaction effects of the audiovisual trust experiment imply that a larger fWHR is generally regarded as being more trustworthy in dating circumstances, but that this effect may also differ between genders. With regard to pitch, it may thus be better to speak in a higher pitch in courtship situations, as this is suggested to be perceived as more trustworthy.

It was hypothesized that lower voices would be perceived as more trustworthy in neutral, but not mating-related contexts. This hypothesis was only partly supported, as lower voices were perceived as less trustworthy for both men and women in both contexts. However, this negative effect becomes more pronounced in mating-related contexts. Some of the hypothesized interaction effects were also confirmed. Most notably the interaction between the content of the sentence, the fWHR, and the gender of the speaker, meaning that the face with the fWHR that is seen as most trustworthy depends on the context it is placed in, and that this effect differs between men and women. Another interesting interaction effect is between the content of the sentence and vocal pitch, which was also found. The fWHR was, however, not found to interact with the content of the sentence, nor gender. Evidence was also found in support of the hypothesized interaction between vocal pitch and gender of the

speaker, suggesting that pitch might work as a trustworthiness cue reliant on the gender of the speaker.

Comparison of unimodal trust and attractiveness scores revealed that low-pitched male voices are perceived as more attractive, but at the same time as less trustworthy in mating-related contexts. This perhaps paradoxical finding provides interesting insights into mate preferences based on the voice, and is consistent with previous research suggesting that low-pitched male voices are more attractive (Feinberg, 2005), but inconsistent with research suggesting lower-pitched voices are seen as more trustworthy (Schirmer, 2020). This inconsistency may stem from the difference between long-term and short-term mating intentions, as proposed by previous research (Puts, 2005). Nonetheless, this finding emphasizes the ongoing debate of acoustical parameters and trustworthiness (Belin, 2017; Schirmer, 2020). The fWHR was an important factor for both trustworthiness and attractiveness. It could therefore be the case that the face is seen as trustworthy if it is attractive, or the other way round. This theorization could not be generalized to the voice, which may suggest that trustworthiness judgments rely on different, yet cooperative mechanisms for the face and the voice.

A clear limitation of this experiment is its subject pool. As stated before, this experiment was presented to a small number of participants (N=33), the majority of which being women (N=26). This will no doubt have had an effect on the data, and might even explain why male voices were rated as more attractive as compared to female voices. For study 2, this problem is more prominent, as this experiment was presented to only about half of the participants, as indicated before. Additionally, there were too few male participants to analyse differences in male and female listeners, which may have produced more information.

Another issue for this study is that participants were given no clear definition of trust to abide by for the duration of the experiment. This implies that participants are likely to have adhered to their personal concept of trust, and this may differ between subjects. For instance, it might

be the case that the trustworthiness of a voice or face is interpreted differently if part of one's definition of trust includes competence instead of harmlessness. It is advisable for future studies to take this into consideration.

Lastly, it may not have been clear whether the mating-related sentences hinted at long-term or short-term relationships. Research has shown that mate preferences for both men and women differ for the intention of the exchange (Jonason, 2013). Sentences used in this study may fall more towards short-term relationships, but this remains somewhat ambiguous. Future research may consider taking this into consideration, as it may cause results to differ. Nevertheless, this experiment provided evidence to support claims made by previous research (Barclay, 2017; Schild, 2019; Tsankova, 2015), as well as deepened knowledge on the vocal and facial cues that influence trust and attractiveness across neutral and mating-related contexts. It also shows that there are many influences at play in daily conversations and that these influences interact. This highlights the complexity of this type of judgment

In conclusion, the multimodality of human trust perception is highly dependent on the context of the interaction, and is mediated by the facial features and vocal pitch of the speaker. People seem to be more careful in trusting another in dating circumstances. Further, a lower-pitched voice is generally perceived as less trustworthy in mating-related, but not neutral, contexts. Finally, these effects could not be explained through attractiveness, but it has been found that lower-pitched male voices are simultaneously seen as more attractive yet less trustworthy. Given the complexity of audiovisual trust judgments demonstrated in this study, more research on the multimodality of human trust interaction is required in order to form a conclusive theory.

References

Adobe Photoshop CS. (2004). Berkeley, CA: Peachpit Press.

Alaei, Ravin, et al. (2020). Accuracy and Bias in First Impressions of Attachment Style From Faces. *Journal of Personality, 88(5)*, 940-949. https://doi.org/10.1111/jopy.12540

Barclay, P. & O'Connor, J. J. M. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior, 38(4),* 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001

Belin P, Boehme B, McAleer P (2017) The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS ONE 12(10)*. e0185651. https://doi.org/10.1371/journal.pone.0185651

Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 5.3.51. 2013. Available: http://www.praat.org/

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour 69(3),* 561–568. https://doi.org/10.1016/j.anbehav.2004.06.012

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance, 40(4),* 1526–1541. https://doi.org/10.1037/a0036732

Geniole, S. N., Carré, J. M., Denson, T. F., Doxson, B. J., & McCormick, C. M. (2015). Evidence from Meta-Analyses of the Facial Width-to-Height Ratio as an Evolved Cue of Threat. *PLoS ONE 10(7):*e0132726 https://doi.org/10.1371/journal.pone.0132726

Haselhuhn, M. P.,Ormiston, M.E., & Wong, E.M. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences 105.* 40-42. https://doi.org/10.1016/j.paid.2016.09.017

Hirschberg, J, Levitan, S.I., & Maredia, A. (2018). Linguistic Cues to Deception and Perceived Deception in Interview Dialogues. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* 1941–1950. Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1176

Jonason, P. K., Webster, G. D., & Gesselman, A. N. (2013). The Structure and Content of Long-Term and Short-Term Mate Preferences. *Interpersona: An International Journal on Personal Relationships, 7(2),* 167–179. https://doi.org/10.5964/ijpr.v7i2.125

Kramer, R.S. (2017). Sexual dimorphism of facial width-to-height ratio in human skulls and faces: A meta-analytical approach. *Evolution and Human Behavior, 38*, 414-420. https://doi.org/10.1016/j.evolhumbehav.2016.12.002

Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). Facial attractiveness: evolutionary based research. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 366(1571).* 1638–1659. https://doi.org/10.1098/rstb.2010.0404

Mathôt, S. S. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314-324.

Merlhiot, G., Mondillon, L., Méot, A., Dutheil, F., & Mermillod, M. (2021). Facial width-to-height ratio underlies perceived dominance on facial emotional expressions. *Personality and Individual Differences 172.* 110583. https://doi.org/10.1016/j.paid.2020.110583

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior 26(5).* 388-397. https://doi.org/10.1016/j.evolhumbehav.2005.03.001

Schild, C., Stern, J., & Zettler, I. (2019). Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology 31(1),*164-175. https://doi.org/10.5061/dryad.q2db26m

Schirmer A, Chiu M. H., Lo, C., Feng Y-J, & Penney T. B. (2020) Angry, old, male – and trustworthy? How expressive and person voice characteristics shape listener trust. *PLoS ONE 15(5).* e0232431. https://doi.org/10.1371/journal.pone.0232431

Stirrat M & Perrett D.I. (2010). Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychological Science 21(3),* 349-354. https://doi.org/10.1177/0956797610362647

Suire, A., Raymond, M., & Defradas, M. (2019). Male Vocal Quality and Its Relation to Females' Preferences. *Evolutionary Psychology. 17(3):147470491987467* https://doi.org/10.1177/1474704919874675

Tsankova, E., Krumhuber, E., Aubrey, A. J., Kappas, A., Möllering, G., Marshall, D., & Rosin, P. L. (2015). The multi-modal nature of trustworthiness perception. *Proceedings of the 1st Joint Conference on Facial Analysis, Animation, and Auditory-Visual Speech Processing* (FAAVSP). 147-152. ISCA: Vienna, Austria. https://discovery.ucl.ac.uk/id/eprint/1475189

Ward, J. (2016). *The Student's Guide to Social Neuroscience* (2nd ed.). Milton Park: Taylor & Francis Ltd.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science, 26(3),* 237–242. https://doi.org/10.1177/0963721416683996