RACE, SEX, VOCAL CHARACTERISTICS AND EMOTION AFFECT TRUST OF AUDITORY WITNESS TESTIMONY

THE INTERSECTIONALITY OF SEX, RACE, VOICE PITCH, AND EMOTION ON PERCEPTIONS OF TRUST OF AUDITORY WITNESS TESTIMONY

By

CHARLENE G. V. FORDE-SMITH

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For a degree

Master of Science

McMaster University

MASTER OF SCIENCE (2023) McMaster University

(Psychology, Neuroscience & Behaviour) Hamilton, Ontario

TITLE: The Intersectionality of Sex, Race, Vocal Characteristics and Emotion on Perceptions of Trust of Auditory Witness Testimony

AUTHOR: Charlene Forde-Smith BA (McMaster University)

SUPERVISOR: Dr. David R. Feinberg

Number of Pages: 49

Abstract

Trust is valuable as it plays a vital role in first impressions and decision-making. However, trust perceptions of speakers are heavily influenced by stereotypes and biases. Given how impactful eyewitness testimonies are in courtroom rulings and how often biases are used to judge speakers in courtroom settings, it is crucial to understand what factors impact perceptions of trust within this context. This is the first study to analyze the relationship between trust perception and emotion (Anger, Fear, Happy, Sad, Disgust, and Neutral) within the context of a courtroom testimony while also looking at how race, sex, vocal characteristics of the speaker, and intensity (gun-present vs. gun-absent crime) impact this interaction. Participants listened to a random sample of voices saying, "That is exactly what happened" and then responded yes or no when asked if they trusted the speaker.

We found a highly significant interaction between sex and race on the proportion of voices trusted in select emotions. An in-depth analysis of voice characteristics indicated varying effects of pitch, Cepstral Peak Prominence, Vocal tract length, Subharmonic to harmonic ratio, Speech rate, Long-term Average Spectrum, and Harmonics to Noise Ratio (HNR) on perceptions of trust in male and female speakers.

This experiment supports findings that heuristic cues influence the perception of trust in the courtroom. Understanding the role stereotypes and biases play in decision-making in the courtroom is vital to ensuring a fair prosecution.

Acknowledgements

I would like to thank my supervisor, David Feinberg, for his support and guidance throughout my time in the voice lab. I would also like to thank my family, Charmaine, David, and Daveon and my friends for always encouraging me and believing in me throughout my master's.

Material Prepared for Publication

The manuscript in Chapter 2: "Race, sex, and emotion affect trust of auditory witness testimony" is adapted from a manuscript prepared for publication:

Forde-Smith, G. V. C., & Feinberg, D. R. (2023). Race, sex, and emotion affect trust of auditory witness testimony. (prepared).

Author Roles:

Forde-Smith, G.V. C: Voice manipulations, experimental design, data collection, data analysis, wrote paper.
Feinberg, D.R.: Experimental design, data collection, data analysis, theoretical supervisor, edited paper.

Table of Contents

List of Tables and Figures

Chapter 1: General Introduction

1.1 Trust

Trust is essential in complex decision-making (Edelenbos & Klijin, 2007). With abundant knowledge gained across disciplines such as psychology and sociology, trust has become increasingly difficult to quantify, interpret, and operationalize (Simpson, 2007). However, trust requires vulnerability as it involves forming expectations about an individual's behaviour (Simpson, 2007). This can make trust challenging to ascertain, and yet, it is still so valuable as it facilitates, solidifies, and enhances cooperation (Edelenbos & Klijin, 2007). Alongside traits like dominance and attractiveness, trust plays a prominent role in first impressions (McAleer et al., 2014). It is especially beneficial when making decisions in situations with a lot of uncertainty and complexity (Edelenbos & Klijin, 2007) and when supporting evidence may be incomplete (Cho et al., 2015). Human motivations, emotions (Simpson, 2007), and heuristic cues (Siegrist, 2021) impact perceptions of trust.

1.2 Trust, race, and sex in the courtroom

Eyewitness testimonies are the most compelling, yet misleading information jurors can receive (Beety, 2022). In fact, "between seventy-five and eighty-five percent of the convictions overturned by DNA evidence have involved a mistaken eyewitness (and yet) 80,000 suspects are targeted every year based on an eyewitness identification (Fedella, 2006). Memory is malleable, and our experiences, beliefs, and personal biases impact our perception and judgement of others (Beety, 2022), leading to misidentifications and inaccurate recounting of crimes. As such, we wanted to see how individuals would perceive audio testimonies from an eyewitness that vary in sex, race, emotion, and acoustic properties of the voice.

Studies have found that race and sex are heuristic cues for male and female jurors and that jurors may use rationalization, stereotypes, and what is socially acceptable to make judgements in the courtroom (Bagby et al., 1999; Devine & Caughlin, 2014; Jones & Kaplan et al., 2010; Mitchell et al., 2005; Maeder & Yamamoto, 2015). As such, further understanding how our biases impact decision-making in life-changing situations and drawing the attention of judges and jurors to these biases is critical (Cohn et al., 2009; Bucolo & Cohn, 2010). Judges and jurors are more likely to be biased towards other-raced defendants (Maeder & Yamamoto, 2019). The social identity theory suggests that this is because we are "motivated to maintain favourable views of groups we identify with", which may cause us to trust those who are most like us more than those who are not (Maeder & Yamamoto, 2019). Mock jurors also have biases towards marginalized groups and are harsher towards black (Maedar & Burdett, 2013) and indigenous defendants (FosterLee et al., 2006; Maeder & Yamamoto, 2019). These biases can be explained by the marginalization of and internalized resentment towards minority groups (Mitchell et al., 2005). This may also be because of stereotypes and mental shortcuts as judgment mechanisms (Maeder & Yamamoto, 2019). Sex discrimination is also present in the courtroom, such that raters perceive female attorneys as less intelligent, capable, and experienced than male attorneys (Hodgson & Pryor, 1984; Maeder & Yamamoto, 2015). Emotion also moderates the perception of male and female speakers in the courtroom. Anger benefited how male attorneys were perceived and negatively impacted how female attorneys were perceived (Salerno et al., 2018). While angry females are perceived as shrill, hysterical, and grating, angry males are perceived as commanding, powerful, and competent (Salerno et al., 2018). Raising awareness of possible biases in the courtroom may help reduce the impact of these biases on judgment decisions (Cohn et al., 2009; Bucolo & Cohn, 2010; Maeder & Yamamoto, 2019).

1.3 Trust, face, and emotion

The relationship between trust perception and emotion in faces has been analyzed in previous experiments (Galinsky et al., 2020; Campellone & Kring, 2013; Tortosa et al., 2013). Perceived trustworthiness and happiness ratings are highly correlated when looking at emotional expression (Galinsky et al., 2020). Faces with larger smiles were perceived as more trustworthy (Galinsky et al., 2020). In an economic games study, participants played a modified version of the trust game where they learned players' behavioural patterns (Campellone & Kring, 2013). While happy faces did not affect participant judgments, participants were less likely to invest in individuals with angry faces (Campellone & Kring, 2013). In similar experiments, happy faces resulted in an increased level of trust compared to angry faces (Tortosa et al., 2013).

When considering legal sentencing decisions, defendants with untrustworthy faces were likelier to be perceived as guilty (Jaegar et al., 2020). Defendants with less trustworthy faces are also more likely to receive a death sentence (Wilson & Rule, 2016). This is concerning as individuals are likelier to trust individuals who resemble themselves (Nakano & Yamamoto, 2022). Because of repeated exposure, individuals are more likely to respond positively to faces that they are more familiar with (Nakano & Yamamoto, 2022). Similarly, familiarity makes average face composites more attractive (Rhodes et al., 2001). If individuals are more likely to trust those like them, this can lead to biased judgments of witnesses and defendants in the courtroom. The influence of faces on trust perceptions and decision-making is present in young children, indicating that biases are established early (Ewing et al., 2015).

The persistent reliance on facial trustworthiness is better explained by the intuitive accessibility of facial cues rather than their ability to accurately identify trustworthy vs.

untrustworthy individuals (Campellone & Kring, 2013), which is also the case in voice pitch research (Schild et al., 2020). Face research indicates that trust perception is influenced by voice pitch and participant judgement and decision-making (behaviour). Faces and voices are processed similarly (McAleer et al., 2014), which raises the question, is trust perceived similarly in faces and voices? As emotions are conveyed through vocal expressions (CITE) and are cross-culturally recognized (Sauter et al., 2010; Scherer et al., 2001), we wanted to see if findings would replicate with voice pitch.

1.4 Voice Pitch & Trust

Judgements are made about others within seconds of contact (Bergmann et al., 2012). While visual appearance is a salient cue, we also draw conclusions from non-verbal vocal properties such as voice pitch. Voice pitch is the perception of fundamental frequencies (F0) (Moore, 1995). Pitch is filtered through the vocal tract to produce speech output. Voice pitch influences how high or low a voice is in both male and female speakers (Moore, 1995). Male voice pitch typically ranges between 60 Hz to 180 Hz, while female voice pitch ranges from 160 Hz to 300 Hz (Re et al., 2012). The influence of voice pitch on trust perception is context-dependent (Montano et al., 2017) and can be crucial to the trajectory of the relationship between individuals (Bergmann et al., 2012).

Previous research has looked at the relationship between voice pitch and trust within the context of mating, economic games, voting behaviour, and workplace prestige. In mating studies looking at perceptions of sexual infidelity and preferences for short and long-term relationships, female raters perceived lower-pitched male speakers to be at a higher risk of infidelity, and male raters perceived high-pitched female speakers to be at a high risk of infidelity (O'Connor et al.,

2011). While lower-pitched male speakers were also perceived as more attractive (Feinberg et al., 2005; Schild et al., 2021) in short-term and long-term relationships, they were also perceived as more likely to cheat than their higher-pitched counterparts (O'Connor et al., 2014; Schild et a., 2019). Lower-pitched male speakers also have higher reproductive success and produce more offspring in hunter-gatherer populations (Apicella et al., 2007). Male and female speakers modulate their voice pitch to appear more desirable when interacting with members of the opposite sex that they are attracted to (Pisanski et al., 2018). Preferences for lower-pitched males in mating experiments may point towards a preference for higher testosterone and masculinity levels in males (Feinberg, 2005).

In economic games, a positive correlation exists between voice pitch and trust, such that higher-pitched male speakers were trusted significantly more than lower-pitched males (Montano et al., 2017). Lower-pitched male speakers are also more likely to be trusted to oversee larger firms, receive more compensation, and are retained longer (Mayew et al., 2013). When analyzing voting behaviour, lower-pitched females were perceived as more trustworthy and received more votes than higher-pitched female speakers (Klofstad et al., 2012). Lower-pitched male speakers were perceived as less trustworthy than higher-pitched male speakers (Klofstad et al., 2012); however, they still received more votes (Klofstad et al., 2012; Klofstad et al., 2015). Older, well-educated, and politically engaged voters were also more likely to vote for lower-pitched speakers (Klofstad et al., 2015). As voice pitch preferences vary between contexts and preferences influence our interactions with others, it would be valuable to analyze how they influence perceptions of trust in the courtroom.

1.5 Additional Voice Characteristics

We also explored additional voice characteristics to obtain a more holistic view of their impact on trust perceptions. Cepstral Peak Prominence (CPP) indicates how much noise is in the audio recording while controlling for pitch. Lower CPP indicates a more dysphonic (hoarse) voice (Heman-Ackah et al., 2003; Fraile & Godino-Llorente, 2014). CPP also influences what is considered a good voice quality (Warhurst et al., 2017). Compared to measures of jitter, shimmer, and noise-to-harmonic ratios, Cepstral Peak Prominence is a better measure for dysphonia (Heman-Ackah et al., 2003; Fraile & Godino-Llorente, 2014).

Vocal tract length (VTL) is an estimate of vocal tract length based on formant frequencies (resonant vocal tract frequencies) and affects trust judgments (Belin et al., 2017). Longer vocal tract lengths are perceptually associated with increased masculinity and body size (Feinberg et al., 2005). In males, low formants, often associated with increased masculinity, were perceived as more cooperative than high formants, typically associated with increased femininity (Feinberg et al., 2005; Knowles & Little, 2014). In an economic games study, shortened vocal tract lengths were less likely to be trusted than lengthened formants (Montano et al., 2017). Like voice pitch, vocal tract length and formants are linked to perceptions of social traits (Feinberg et al., 2005).

Subharmonic to harmonic ratio (SHR) is noise related to how creaky a voice sounds (Sun et al., 2002; Imhof et al., 2014). There is a strong relationship between SHR and Pitch. SHR can be present at any voice pitch level, and at certain levels (Imhof et al., 2014), it can alter the perception of voice pitch (Sun et al., 2002). Much like CPP, SHR is often used to analyze the voice quality of vocalizers (Leong et al., 2013)

Speech rate refers to the speed at which vocalizers talk and can impact how much we trust speakers (Miller et al., 1976). Individuals with a faster speech rate were perceived as more credible and persuasive than slower-speaking individuals (Miller et al., 1976). However, faster speech rates negatively impacted the amount of money received in crowdfunding performance. Speech rate also impacts initial perceptions of trust (Velner et al., 2021). Speech rate is perceived differently depending on the context.

Harmonics to Noise Ratio (HNR), is a signal-to-noise pitch measure (Belin et al., 2017). Speakers perceived as more trustworthy typically have lower HNR, while speakers perceived as less trustworthy have a higher HNR (Schirmer et al., 2019). In some studies, younger individuals have lower HNR than older individuals (Schirmer et al., 2019); in other studies, the opposite has been found (Ferrand, 2002). HNR could be used to estimate speaker age (Schirmer et al., 2019). HNR also moderates the perception of human emotion (Schirmer et al., 2019).

Spectral tilt indicates breathiness and predicts voice attractiveness (Xu et al., 2013) which could affect trust ratings. Recent studies have shown that listeners are sensitive to differences in spectral tilt and can use this vocal characteristic to differentiate between ethnic groups (Szakay, 2012). We measured spectral tilt across a long-term average spectrum (LTAS) (O'Connor et al., 2017).

1.6 Importance

Research analyzing the intersectionality between race, sex and emotion on trust perception is important as it highlights judgement biases that may occur in the courtroom. Bringing further awareness to biases in the courtroom is a crucial step when working towards

ensuring that individuals are given an equal prosecution. As eyewitnesses play a significant role in swaying the decisions of judges and jurors in the courtroom, it is important to understand the role that biases and stereotypes have in how they are perceived by others. Bringing further awareness towards how deep our biases run will hopefully allow judges and jurors to be more cautious when making decisions that have lifelong effects.

Chapter 2: Experiment (Manuscript)

2.1 Introduction

2% to 10% of the US prison population are wrongfully convicted of a crime they did not commit, as per an extrapolation conducted by the National Registry of Exonerations (How many innocent people are in prison, 2023). New evidence has exonerated Eight hundred and fifty individuals between 2000 and 2011 (How many innocent people are in prison, 2023). Eyewitness testimonies are the most compelling yet misleading sources of information that jurors can receive (Beety, 2022). In fact, "Between seventy-five and eighty-five percent of the convictions overturned by DNA evidence have involved a mistaken eyewitness (and yet) 80,000 suspects are targeted every year based on an eyewitness identification" (Fradella, 2006). Memory is malleable, and our experiences, beliefs, and personal biases impact how we perceive and judge others (Beety, 2022). Trusting the wrong witness can lead to the misidentification of suspects and inaccurate recounting of crimes by witnesses.

Trust is a complex phenomenon that is difficult to quantify yet valuable to achieve (Cho et al., 2015; Edelenbos & Klijin, 2007; Jaegar et al., 2020; Simpson, 2007). Perceived trustworthiness can impact decision-making and the trajectory of relationships (Galinsky et al., 2020) between individuals in economic games (Campellone & Kring, 2013; Montano et al., 2017), the courtroom (Jaegar et al., 2020; ForsterLee, 2006; Maeder & Yamamoto, 2019), voting scenarios (Klofstad et al., 2012; Klofstad et al., 2012; Tigue et al., 2011), the workplace (Mayew et al., 2013) and in romantic relationships (O'Connor et al., 2011; O'Connor et al., 2014). Whom we trust is affected by our biases towards race (Smith, 2010; `, 2019), sex (Maeder & Yamamoto, 2019), voice characteristics (O'Connor et al., 2011; O'Connor et al., 2014; Montano et al., 2017; Klofstad et al., 2012) and the situation.

When playing an economic trust game, angry individuals were less likely to be trusted than happy individuals when there was an established pattern of behaviour (Campellone & Kring, 2013). There is a positive correlation between happiness and perceived trustworthiness when measuring the degree of smiling in photos (Galinsky et al., 2020). As emotions can be conveyed through vocal characteristics (Protopapas & Lieberman, 1997), much like in faces (De Gelder & Vroomen et al., 2000; Matsumoto et al., 2008) and as basic emotions (anger, fear, disgust, happiness, sadness, and surprise) in voices are to some degree recognized cross-culturally (Sauter et al., 2010; Scherer et al., 2001) we tested the effect of voice pitch and trust within the context of a courtroom.

Humans conclude information about others from non-verbal vocal properties such as voice pitch. In mating studies, raters perceived high-pitched females and low-pitched males as having a higher risk of infidelity (O'Connor et al., 2011). In economic games (Montano et al., 2017), voting (Klofstad et al., 2012), and mating (O'Connor et al., 2011; O'Connor et al., 2014) studies, higher-pitched male speakers were trusted more than lower-pitched voices. High-pitched female speakers are trusted less in voting (Klofstad et al., 2012) and mating (O'Connor et al., 2011) scenarios.

We also explored which acoustic properties of the voice predicted trust ratings and how speaker sex, race, and emotion moderate this relationship. We analyzed the following measures because they affect perceptions of trust and related features and give a well-rounded representation of vocal characteristics. Cepstral Peak Prominence (CPP) indicates how much noise is in the audio recording while controlling for pitch. Lower CPP is indicative of a more dysphonic (hoarse) voice (Heman-Ackah et al., 2003) and predicts trust (Warhurst et al., 2017). Vocal tract length (VTL) is an estimate of vocal tract length based on formant frequencies

(resonant vocal tract frequencies) and affects trust judgments (Belin et al., 2017; Montano et al., 2017). Subharmonic to harmonic ratio (SHR) is noise related to how creaky a voice sounds (Sun et al., 2002; Imhof et al., 2014). Speech rate refers to the speed at which vocalizers talk and can impact how much we trust speakers (Miller et al., 1976). Harmonics to Noise Ratio (HNR), a signal-to-noise pitch measure, also affects trust judgments (Belin et al., 2017). Spectral tilt indicates breathiness and predicts voice attractiveness (Xu et al., 2013) which could affect trust ratings. We measured spectral tilt across a long-term average spectrum (LTAS) (O'Connor et al., 2017).

Eyewitnesses play a significant role in swaying the decisions of judges and jurors. Eyewitness characteristics like race and sex act as heuristic cues for judges and jurors in the courtroom (Maeder & Yamamoto, 2019). As voice pitch and emotion also impact decision-making, we analyzed the relationship between speaker race, sex, vocal characteristics, and emotion of speakers on perceptions of guilt and innocence within a courtroom setting. We predicted that trust perceptions would vary depending on vocalizer emotions and that sex and race would moderate this relationship. As lower-pitched males and higher-pitched females were perceived as less trustworthy in voting scenarios (Klofstad et al., 2015; Tigue et al., 2011) and at a higher risk of infidelity in mating scenarios (O'Connor et al., 2011), we anticipated that raters would perceive lower-pitched male voices and higher-pitched female voices as less trustworthy. Alternatively, raters will perceive lower-pitched women as less trustworthy as lower-pitched voices are perceived as more dominant, and dominance negatively correlates with trust (Tigue et al., 2011). We did not have directional predictions for other acoustic features but wanted to explore how they might affect trust measures. We also predicted that participants would perceive male vocalizers as less trustworthy than women. To test our hypotheses, participants are told that they will

listen to individuals testifying about a robbery at a convenience store where a gun was either present or absent. They then listened to 60 male or female speakers saying, “That is exactly what happened” and selected yes or no when asked if they trusted the speaker. A brief demographic questionnaire followed this section.

2.2 Methods

Participants

This study was advertised and administered online on Prolific.co. Participants received 1.88 euros upon completion of the study. 323 (164 males, 139 females, 2 N/A) diverse (Table 2.1) individuals between the ages of 18 and 76 completed this study. Participants must live in Canada, understand English, and be 18 or older to participate. Participants were of varying ethnic and religious groups. MREB 5560 approved all protocols for this study.

Participant Ethnic Group	
Participant race	Number of participants
White	196
Asian	87
mixed	14
Black	9
Latino	5
Indigenous	5
Jewish	2
Muslim	1
N/A	6

Table 2.1 Participant Demographic breakdown

Audio Stimuli

Audio recordings in this experiment are from the Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D), which is pre-validated for emotion recognition (Cao et al., 2014).

There are 20 audio recordings from 10 African American (5 female & 5 male) and 10 White (5 female & 5 male) vocalizers. Prior work has shown that even though this is a small number of vocalizers, it is sufficient to measure trust judgements reliably (Montano et al., 2017). Speakers said, “That is exactly what happened” in 6 different emotions, neutral, sad, angry, disgusted, fearful, and happy (Figure 2.1). In total, we had 120 different audio recordings. We normalized the amplitude of audio recordings to 70dB RMS using VoiceLab Software (Feinberg, 2022; Feinberg & Cook, 2020).

Figure 2.1

Figure 2.1: Stimuli emotions (neutral, sad, angry, disgust, fear and happy) from the Crowd-sourced Emotional Multimodal Actors Dataset. (Photo credits https://managementmania.com/en/six-basic-emotions)

Procedure

The experiment is programmed in PsychoPy (Pierce et al., 2019) and ported into JavaScript (Wirfs-Brock & Eich, 2020) on Pavlovia (Pierce et al., 2019). Participants accessed this experiment via Prolific.co, completed the information and consent form on our university hosted LimeSurvey instance, and were automatically directed to Pavlovia.co to complete the experiment. Participants are randomly assigned to 1 of 4 blocks in the experiment. Blocks included high-intensity female audio recordings, high-intensity male audio recordings, low-intensity female audio recordings, and low-intensity male audio recordings. Participants listened to audio

recordings of all six emotions. Vocalizer sex and condition intensity (gun present/gun absent) were between-subjects variables, and emotion was a within-subjects variable.

Participants completing the low-intensity condition were told that all proceeding audio recordings were from speakers testifying in court about a robbery at a convenience store (gun-absent condition). Participants in the high-intensity condition block were told that all proceeding audio recordings were from speakers testifying in court about a robbery at a convenience store where a gun was present (gun-present condition). Participants then listened to 60 audio recordings of either male or female speakers saying, "That is exactly what happened". After each audio recording, participants responded yes or no when asked if they would trust the speaker. Participants then completed a brief demographic questionnaire before being redirected to LimeSurvey to complete the debriefing and submit the experiment. Figure 1 displays the different experimental conditions.

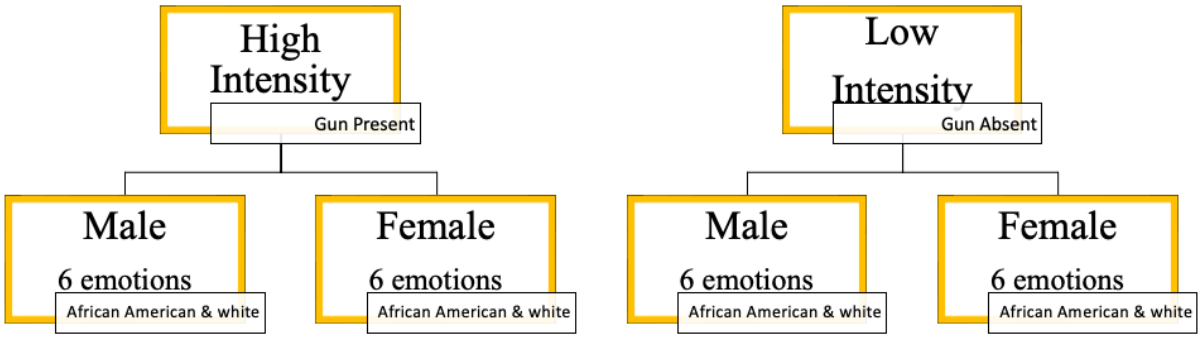


Figure 2.2: Block randomization of raters. Raters participated in one of four blocks. High intensity (gun present) African American and White male speakers displaying all 6 emotions, high intensity (gun present) African American and White female speakers displaying all 6 emotions, low intensity (gun absent) African American and White male speakers displaying all 6 emotions and low intensity (gun absent) African American and White female speakers displaying all 6 emotions.

Statistical model

We used a generalized linear mixed effects model with binomial logit link functions. All binary variables were sum-to-zero coded (-0.5, 0.5), including Trust, such that positive values indicate that a voice was trusted, and negative values indicate distrusted voices. Voice pitch was z-scored. We specified a random slope for each within-subject association of race of voice by rater (Barr et al., 2013; Barr, 2013). Running a model with all emotions became computationally impractical and uninterpretable. Thus, we ran the model separately for each emotion. The summarized results are in Table 1. The complete model, code, and output are in the supplementary materials https://osf.io/mfwye/ . We specified random intercepts for participants and vocalizers to control for interdependence among voices and raters.

2.3 Results

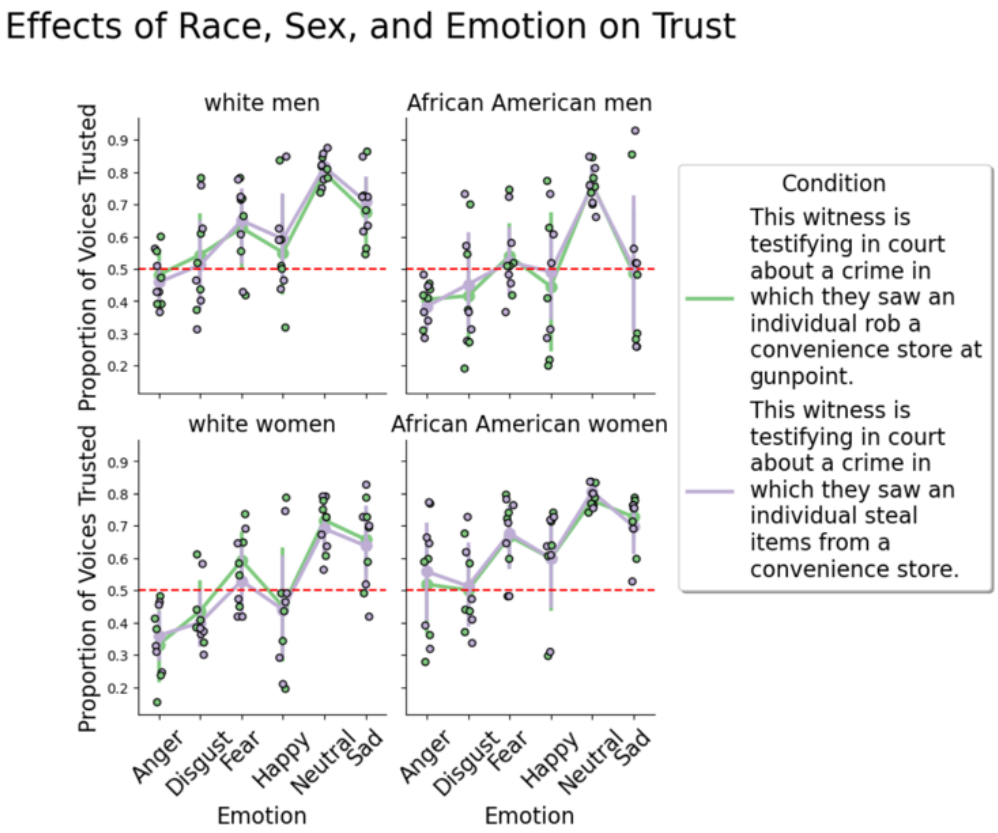


Figure 2.3: Interaction between emotion (x-axis) and proportion of voices trusted (y-axis) separated by speaker sex, race, and condition intensity: the gun present, high intensity condition

is depicted in green, and the gun absent, low intensity condition is depicted as a purple line. the red line represents chance.

Upon visual inspection of Figure 2.3, we found that there was no significant difference (all error bars overlapped strongly) or visible difference between high (gun present) and low (gun absent) intensity conditions. We combined both conditions for all subsequent analyses (Figure 2.3).

A larger number indicates a higher probability of being trusted if the value is positive or less trusted if the value is negative (Table 2.2). The asterisks indicate that a value significantly differs from zero (* $p<0.05$; ** $p<0.01$; *** $p<0.001$). uci represents the upper confidence interval and lci represents the lower confidence interval. There was a highly significant positive interaction between speaker sex and race on the proportion of voices trusted such that speakers who sounded angry (Estimate= 1.44, lci=0.36, uci=2.52, p=0.009), fearful (Estimate=1.06, lci=0.05 , uci=2.07, p=0.039), and neutral (Estimate=0.89, lci=0.32 , uci=1.45 , p=0.002) were significantly likely to be trusted. White males were trusted the most, followed by African American females, White females, and lastly African American males (Figure 2.4). There were no significant main effects of either Sex or Race on perceptions of trust.

Table 2.2: Generalized linear mixed effects model for emotion

	Anger	Disgust	Fear	Happy	Neutral	Sad
Intercept	-0.36 * [-0.65, -0.08]	-0.15 [-0.45, 0.16]	0.51 *** [0.25, 0.78]	0.10 [-0.31, 0.51]	1.58 *** [1.40, 1.77]	0.73 *** [0.38, 1.08]
Sex	0.05 [-0.53, 0.62]	0.12 [-0.50, 0.73]	-0.14 [-0.67, 0.38]	0.05 [-0.78, 0.88]	0.26 [-0.10, 0.62]	-0.40 [-1.10, 0.31]

Race	-0.36 [-0.90, 0.19]	0.04 [-0.56, 0.63]	0.06 [-0.45, 0.56]	-0.09 [-0.90, 0.73]	-0.13 [-0.43, 0.16]	0.33 [-0.36, 1.02]
Sex:Race	1.46 ** [0.37, 2.54]	0.89 [-0.31, 2.08]	1.05 * [0.03, 2.06]	1.32 [-0.30, 2.95]	0.89 ** [0.33, 1.45]	1.27 [-0.11, 2.65]
*** p < 0.001; ** p < 0.01; * p < 0.05						95% [lower ci, upper ci]

Table 2.2: Generalized linear mixed effects model fit by the maximum likelihood (Laplace Approximation). Large negative numbers indicate a higher probability of being trusted. Large negative numbers indicate a higher probability of being perceived as less trustworthy. The asterisks indicate that a value is significantly different from zero. Values in the brackets are confidence intervals.

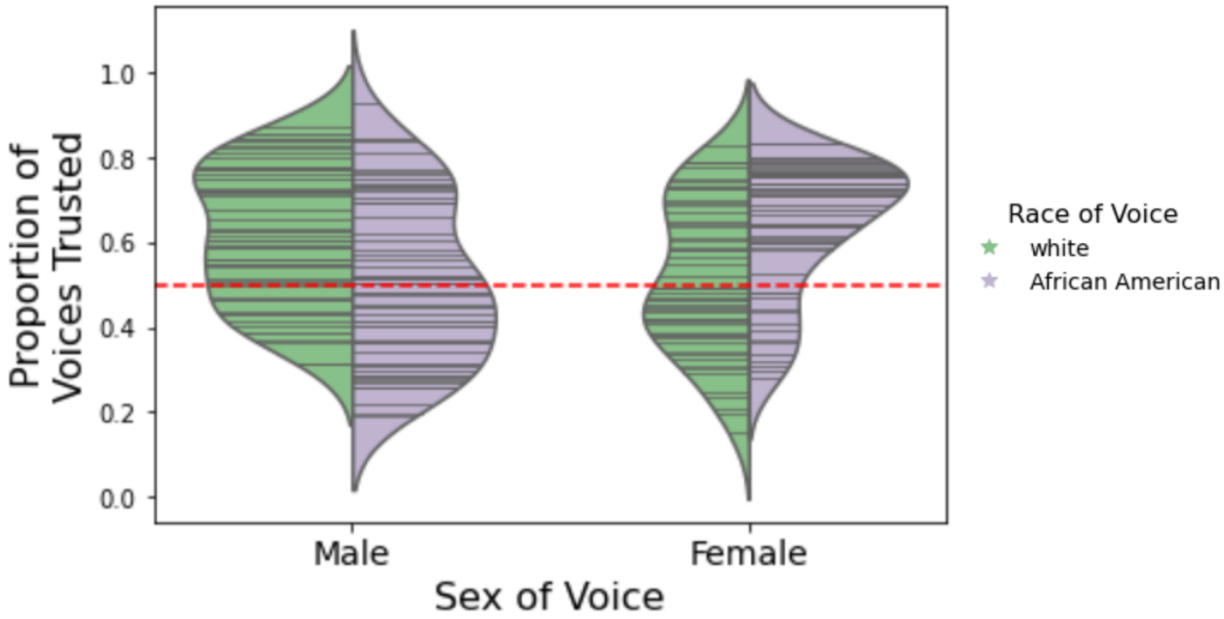


Figure 2.4: Interaction between speaker sex (x-axis) and race on proportion of voices trusted (y-axis). Race of White speakers depicted in green and African American speakers in purple. The red line depicts chance. Higher density of data are depicted at darker points of the graph.

There was a main effect of the intercept, such that speakers who sounded angry (Estimate=-0.35, lci=-0.64 , uci=-0.06 , p=0.017) were significantly less likely to be trusted, and those who sounded fearful (Estimate=0.53, lci=0.26 , uci=0.79 , p=0.000), neutral

(Estimate=1.60, lci=1.41 , uci=1.78 , p=0.000) and sad (Estimate=0.74, lci=0.39 , uci=1.09, p=0.000) were significantly more likely to be trusted. There were no main effects of intercept for happy and disgusted categories, indicating, in general, these voices were neither trusted nor distrusted.

Analysis of Acoustic Properties and Race on Trust

The subsequent analysis explores the effects of acoustic properties and race on trust separately for each sex (Figure 2.5).

Race

African American females are trusted significantly more than White females when sounding angry (Estimate=0.52, lci=0.01, uci=1.02, p=0.045) and sad (Estimate=0.029, lci=0.027, uci=0.030, p=0.000) and are trusted less when sounding disgusted (Estimate=-1.05, lci=-1.40, uci=-0.71, p=0.000). White males are trusted significantly more than African American males when sounding fearful (Estimate=2.00, lci=1.55, uci=2.46, p=0.000) and disgusted (Estimate=0.39, lci=0.12, uci=0.65, p=0.005) and are trusted less when sounding angry (Estimate=-2.59, lci=-3.79, uci=-1.40, p=0.000) (Table 2.3).

Lower-*pitched* male speakers were perceived as more trustworthy when sounding angry (Estimate=-1.51, lci=-2.02 , uci=-1.00 , p=0.000) and less trustworthy when sounding fearful (Estimate=0.76, lci=0.32 , uci=1.21 , p=0.001), happy (Estimate=1.42, lci=0.83, uci=2.01, p=0.000), and neutral (Estimate=1.01, lci=0.47 , uci=1.55 , p=0.000). Lower-pitch female speakers were perceived as more trustworthy when angry (Estimate=-0.54, lci=-1.01, uci=-0.07, p=0.025), happy (Estimate=-4.84, lci=-5.83, uci=-3.85, p=0.000), and sad (Estimate=-1.57, lci=-1.57, uci=-1.57, p=0.000).

Cepstral Peak Prominence (CPP) indicates how much noise is in the audio recording while controlling for pitch. Lower CPP indicates a more dysphonic (horsey) voice (Heman-Ackah et al., 2003). Males with a higher CPP were perceived as more trustworthy when sounding angry (Estimate=0.81, lci=0.51 , uci=1.11 , p=0.000) and disgusted (Estimate=0.57, lci=0.30 , uci=0.83 , p=0.000) and less trustworthy when sounding fearful (Estimate=-0.42, lci=-0.80 , uci=-0.03, p=0.033), happy (Estimate=-9.71, lci=-11.94 , uci=-7.49 , p=0.000) and neural (Estimate=-0.83, lci=-1.32 , uci=-0.34 , p=0.001). Females with a higher CPP were perceived as more trustworthy when sounding sad (Estimate=0.22, lci=0.22 , uci=0.22 , p=0.000) and less trustworthy when sounding angry (Estimate=-0.48, lci=-0.68 , uci=-0.28 , p=0.000).

Vocal tract length (VTL) estimates one's vocal tract length. Female speakers with a higher VTL were perceived as less trustworthy when happy (Estimate=-1.23, lci=-1.63 , uci=-0.83 , p=0.000) and sad (Estimate=-0.51, lci=-0.51 , uci=-0.51 , p=0.000). Males with higher VTLs were perceived as more trustworthy when sounding fearful (Estimate=0.88, lci=0.62 , uci=1.14 , p=0.000) and happy (Estimate=3.44, lci=2.60 , uci=4.28 , p=0.000) and less trustworthy when sounding angry (Estimate=-0.90, lci=-1.24 , uci=-0.56 , p=0.000).

Subharmonic to harmonic ratio (SHR) is noise related to low-pitch voices. Males with higher SHR are perceived as more trustworthy when sounding angry (Estimate=0.46, lci=0.23, uci=0.69, p=0.000), fearful (Estimate=0.13, lci=0.01, uci=0.26, p=0.034), and happy (Estimate=0.73, lci=0.50, uci=0.95, p=0.000), and females with a higher SHR were perceived as more trustworthy when sounding fearful (Estimate=0.51, lci=0.06, uci=0.96, p=0.021) and sad (Estimate=0.65, lci=0.65, uci=0.65, p=0.000).

Speech rate refers to the speed at which vocalizers are talking. Faster male speakers are perceived as more trustworthy when sounding angry (Estimate=1.93, lci=1.12, uci=2.74, p=0.000) disgusted (Estimate=1.09, lci=0.92, uci=1.25, p=0.000), happy (Estimate=2.21, lci=1.68, uci=2.74, p=0.000), and less trustworthy when sounding neutral (Estimate=-0.95, lci=-1.45, uci=-0.45, p=0.000). Faster female speakers are perceived as more trustworthy when happy (Estimate=0.40, lci=0.08, uci=0.72, p=0.015) and less trustworthy when sounding angry (Estimate=-0.91, lci=-1.32, uci=-0.51, p=0.000), disgusted (Estimate=-0.33, lci=-0.65, uci=-0.01, p=0.043), and sad (Estimate=-1.02, lci=-1.02, uci=, p=-1.02).

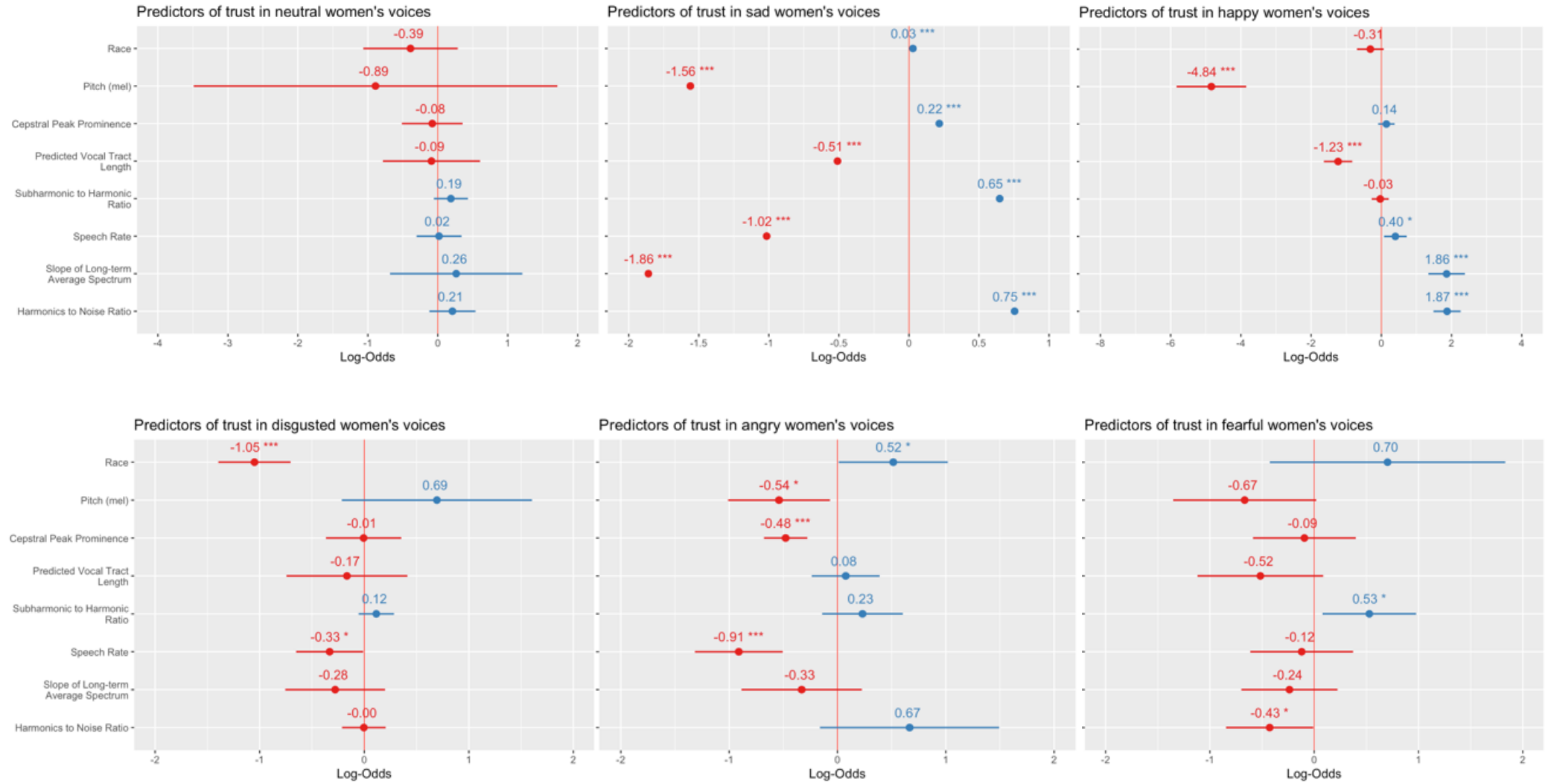
Long-term Average Spectrum (LTAS) is a power spectrum averaged over the entire sound, which indicates how loud low frequencies are compared to high frequencies. Brassier female voices were perceived as more trustworthy when speakers sounded sad (Estimate=-1.87, lci=-1.87, uci=-1.86, p=0.000), while Fluty female speakers were perceived as more trustworthy when happy (Estimate=1.86, lci=1.34, uci=2.38, p=0.000). Fluty male voices were perceived as more trustworthy when speakers sounded angry (Estimate=1.55, lci=0.89, uci=2.21, p=0.000) and happy (Estimate=4.78, lci=3.71, uci=5.85, p=0.000).

Harmonics to Noise Ratio (HNR). A recording with a higher HNR value has more noise. Female speakers with a higher HNR are perceived as more trustworthy when sounding happy (Estimate=1.87, lci=1.49, uci=2.26, p=0.000) and sad (Estimate=0.76, lci=0.76, uci=0.76, p=0.000) and less trustworthy when sounding fearful (Estimate=-0.44, lci=-0.86, uci=-0.02, p=0.045). Male speakers with a higher HNR are perceived as more trustworthy when sounding disgusted (Estimate=0.29, lci=0.09, uci=0.48, p=0.003), fearful (Estimate=0.98, lci=0.66, uci=1.31, p=0.000), happy (Estimate=2.57, lci=1.68, uci=3.45, p=0.000), and less trustworthy when sounding neutral (Estimate=-0.82, lci=-1.27, uci=-0.38, p=0.000).

When testing each acoustic feature as a single covariate in the race * sex model, we found vastly different results than when testing all the acoustic predictors together. We reported the model with acoustic features together here instead of single predictor models as this mirrors our real-world experience more so than does single predictors. Given these discrepancies, we urge researchers to exercise caution when interpreting the nature of the results on acoustic predictors of trust and suggest future replication to elucidate these discrepancies. Data and code are available online here (https://osf.io/mfwye/) for anyone who wishes to explore these data further.

Table 2.3 Generalized Linear Mixed Effects Model of Emotion, Race, and Sex on Trust												
	Anger		Disgust		Fear		Happy		Neutral		Sad	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Intercept	**-0.89*** [-1.16, -0.62]**	0.04 [-0.26, 0.34]	0.01 [-0.16, 0.19]	0.02 [-0.49, 0.52]	-0.24 [-0.50, 0.02]	**1.31 *** [0.71, 1.91]**	**-5.08 *** [-6.70, -3.46]**	**-0.38 *** [-0.59, -0.16]**	**2.49 *** [1.92, 3.06]**	**1.18 * [0.02, 2.35]**	1.54 [-0.37, 3.45]	**0.38 *** [0.38, 0.39]**
Race-effect	**-2.59*** [-3.79, -1.40]**	**0.52* [0.01, 1.02]**	**0.39** [0.12, 0.65]**	**-1.05*** [-1.40, -0.71]**	**2.00*** [1.55, 2.46]**	0.67 [-0.46, 1.81]	0.63 [-0.31, 1.57]	-0.31 [-0.69, 0.07]	-0.39 [-1.00, 0.21]	-0.39 [-1.06, 0.28]	1.57 [-0.36, 3.50]	**0.03 *** [0.03, 0.03]**
pitch	**-1.51*** [-2.02, -1.00]**	**-0.54* [-1.01, -0.07]**	0.07 [-0.16, 0.29]	0.69 [-0.22, 1.60]	**0.76*** [0.32, 1.21]**	-0.64 [-1.33, 0.05]	**1.42 *** [0.83, 2.01]**	**-4.84 *** [-5.83, -3.85]**	**1.01 *** [0.47, 1.55]**	-0.89 [-3.48, 1.70]	-0.20 [-1.25, 0.84]	**-1.57 *** [-1.57, -1.57]**
Cpp	**0.81*** [0.51, 1.11]**	**-0.48*** [-0.68, -0.28]**	**0.57*** [0.30, 0.83]**	-0.01 [-0.37, 0.35]	**-0.42* [-0.80, -0.03]**	-0.07 [-0.57, 0.42]	**-9.71 *** [-11.94, -7.49]**	0.14 [-0.09, 0.38]	**-0.83 *** [-1.32, -0.34]**	-0.08 [-0.51, 0.35]	-0.75 [-4.34, 2.84]	**0.22 *** [0.22, 0.22]**
vtl	**-0.90*** [-1.24, -0.56]**	0.08 [-0.24, 0.39]	0.07 [-0.14, 0.28]	-0.17 [-0.75, 0.41]	**0.88*** [0.62, 1.14]**	-0.53 [-1.13, 0.07]	**3.44 *** [2.60, 4.28]**	**-1.23 *** [-1.63, -0.83]**	0.21 [-0.19, 0.61]	-0.09 [-0.78, 0.60]	-0.99 [-2.61, 0.63]	**-0.51 *** [-0.51, -0.51]**
Subharmonics to harmonics ratio	**0.46*** [0.23, 0.69]**	0.23 [-0.14, 0.60]	0.19 [-0.03, 0.41]	0.11 [-0.05, 0.28]	**0.13* [0.01, 0.26]**	**0.51 * [0.06, 0.96]**	**0.73 *** [0.50, 0.95]**	-0.03 [-0.28, 0.21]	0.21 [-0.03, 0.46]	0.19 [-0.06, 0.43]	-0.96 [-4.90, 2.99]	**0.65 *** [0.65, 0.65]**
speechrate	**1.93*** [1.12, 2.74]**	**-0.91*** [-1.32, -0.51]**	**1.09*** [0.92, 1.25]**	**-0.33* [-0.65, -0.01]**	-0.05 [-0.17, 0.06]	-0.10 [-0.60, 0.39]	**2.21 *** [1.68, 2.74]**	**0.40 * [0.08, 0.72]**	**-0.95 *** [-1.45, -0.45]**	0.02 [-0.30, 0.34]	0.42 [-0.24, 1.08]	**-1.02 *** [-1.02, -1.02]**
LTAS Slope	**1.55*** [0.89, 2.21]**	-0.33 [-0.89, 0.23]	0.06 [-0.13, 0.26]	-0.28 [-0.75, 0.20]	-0.38 [-0.79, 0.02]	-0.23 [-0.69, 0.23]	**4.78 *** [3.71, 5.85]**	**1.86 *** [1.34, 2.38]**	-0.06 [-0.27, 0.16]	0.26 [-0.68, 1.20]	0.91 [-1.48, 3.29]	**-1.87 *** [-1.87, -1.86]**
zHNR	-0.09 [-0.49, 0.32]	0.67 [-0.16, 1.49]	**0.29** [0.09, 0.48]**	-0.00 [-0.21, 0.20]	**0.98*** [0.66, 1.31]**	**-0.44 * [-0.86, -0.02]**	**2.57 *** [1.68, 3.45]**	**1.87 *** [1.49, 2.26]**	**-0.82 *** [-1.27, -0.38]**	0.21 [-0.12, 0.54]	0.62 [-0.42, 1.65]	**0.76 *** [0.76, 0.76]**
*** p < 0.001; ** p < 0.01; * p < 0.05										95% [lower ci, upper ci]		

Table 2.3: Generalized linear mixed effects model fit by the maximum likelihood (Laplace Approximation). Large negative numbers indicate a higher probability of being trusted. Large negative numbers indicate a higher probability of being perceived as less trustworthy. The asterisks indicate that a value is significantly different from zero. Values in the brackets are confidence intervals.


Predictors of trust in neutral women's voices
Race
Pitch (mel)
Cepstral Peak Prominence
Predicted Vocal Tract Length
Subharmonic to Harmonic Ratio
Speech Rate
Slope of Long-term Average Spectrum
Harmonics to Noise Ratio
-0.39
-0.89
-0.08
-0.09
0.19
0.02
0.26
0.21
Log-Odds
Predictors of trust in sad women's voices
0.03 ***
-1.56 ***
0.22 ***
-0.51 ***
0.65 ***
-1.02 ***
-1.86 ***
0.75 ***
Log-Odds
Predictors of trust in happy women's voices
-0.31
-4.84 ***
0.14
-1.23 ***
-0.03
0.40 *
1.86 ***
1.87 ***
Log-Odds
Predictors of trust in disgusted women's voices
-1.05 ***
0.69
-0.01
-0.17
0.12
-0.33 *
-0.28
-0.00
Log-Odds
Predictors of trust in angry women's voices
0.52 *
-0.54 *
-0.48 ***
0.08
0.23
-0.91 ***
-0.33
0.67
Log-Odds
Predictors of trust in fearful women's voices
0.70
-0.67
-0.09
-0.52
0.53 *
-0.12
-0.24
-0.43 *
Log-Odds

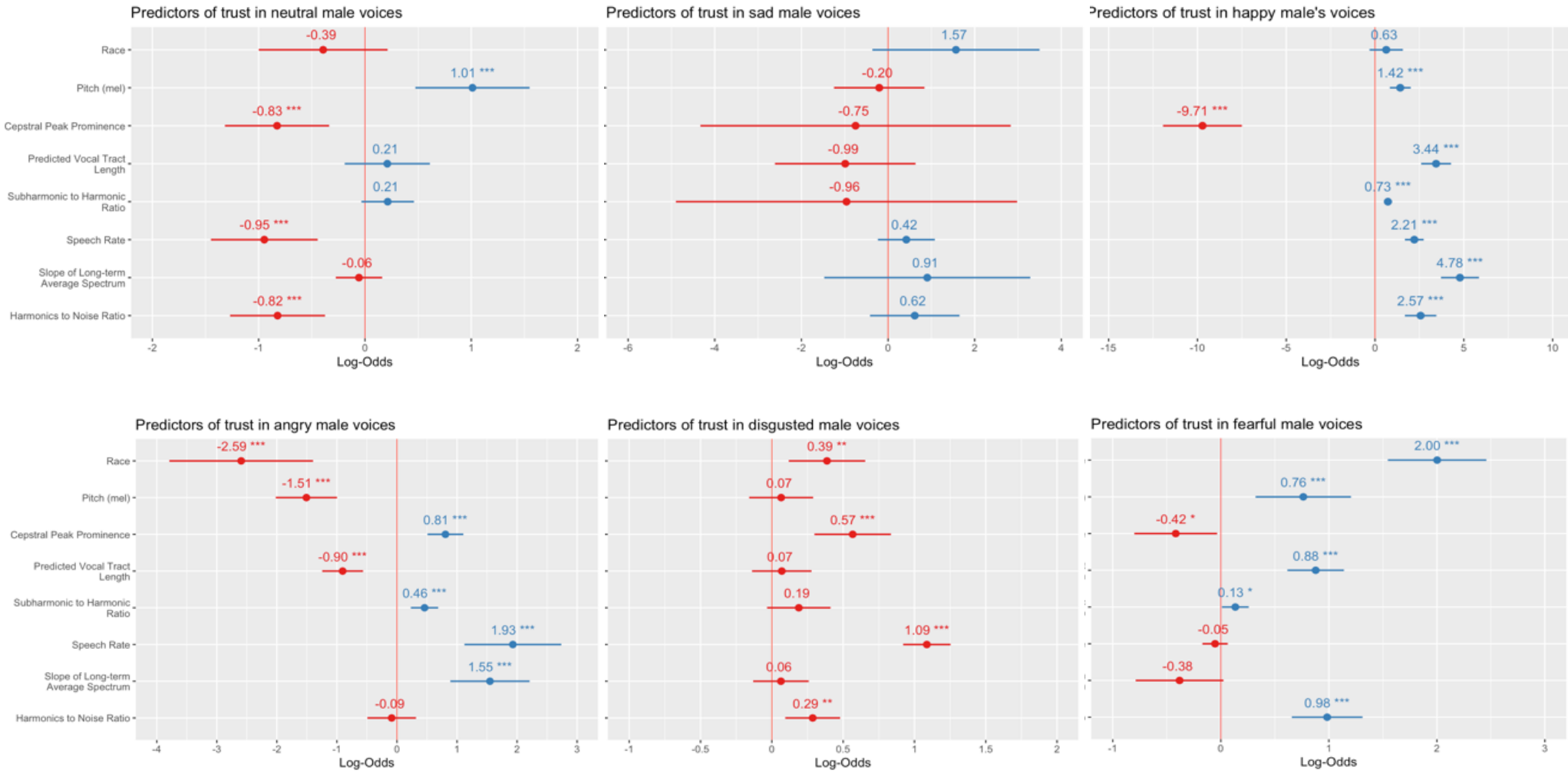


Figure 2.5: Effects of acoustic properties and race on trust of male speakers separated by emotion. A large number is indicative of a higher probability of being trusted (blue), if the value is positive, or being less trusted, if the value is negative (red). the asterisks indicate that a value is significantly different from zero. the red line indicates neutral trust perceptions.

2. 4 Discussion

Effect of Sex & Race

We found a significant interaction between speaker sex and race when speakers were angry, fearful, and neutral but not disgusted or sad. Interestingly, on their own, sex and race did not affect trust, but when put together, White males were trusted the most, followed by African American females, White females, and African American males. While studies have found that race is a heuristic cue for judges and jurors in the courtroom, research has focused primarily on this interaction for defendants (Maeder & Yamamoto, 2019). Jurors may use stereotypes or mental shortcuts to make decisions (Maeder & Yamamoto, 2019); however, further research is needed to examine eyewitnesses. Our experiences, beliefs, and personal biases impact our perception and judgement of others (Beety, 2022). African American males were trusted the least, which supports prior findings that negative stereotypes of African American Men have a life-altering impact on the outcome of courtroom trials (Maeder & Yamamoto, 2019). While White males were trusted more, this trustworthiness rating has no relation to these speakers' actual trustworthiness (validity), but rather biases rooted in racism and colonialism (Chan & Cardenas, 2021). We hypothesize that White Women were less likely to be trusted compared to other groups due to the term deemed "the Karen Effect," which, as described by Times magazine, relates to the attention being drawn to the exploitation of privileges "following a long and troubling legacy of White women in the country weaponizing their victimhood" (Lang, 2020).

Emotion

Lower-pitched male and female speakers were trusted more than higher-pitched male and female speakers. While our findings do support the voice pitch literature that states that raters perceived lower-pitched female speakers as more trustworthy, they do not support voting (Klofstad et al., 2012), economic games (Montano et al., 2017), or mating (O'Connor et al., 2011) literature that found that raters perceived lower pitched male speakers as less trustworthy. Context moderates the interaction between speaker sex and voice pitch.

Effect of Sex & Voice Measurements

Lower pitched male and female speakers were trusted more than higher pitched male and female speakers. While our findings do support the voice pitch literature that states that lower pitched female speakers are perceived as more trustworthy, they do not support the voting (Klofstad et al., 2012), economic games (Montano et al., 2017) or mating (O'Connor et al., 2011) literature that found that lower pitched male speakers were perceived as less trustworthy. As such, the interaction between speaker sex and voice pitch is moderated by context.

Cepstral peak prominence indicated that noisier angry voices were trusted more, and less noisy sad voices were trusted more. Happy and sad voices with lower estimated vocal tract lengths were trusted more. Vocal tracts are shortened when smiling (Ohala, 1983) which may help listeners identify genuine emotions. Fearful and sad voices with louder subharmonics, indicating less creaky voices were more trustworthy in those contexts. Slower speaking, angry, disgusted, and sad voices were trusted more, but faster, happy voices were also trusted more. More breathy happy voices and less breathy sad voices indicated by the LTAS were trusted more. Finally, less noisy happy and sad voices and more noisy fearful voices were trusted more. We reported the results from analyses when all acoustic features were together as covariates.

Analysis of single voice features indicated several directional changes from what was reported above, so it is difficult to say anything conclusive about these acoustic features and how they predict trust. Therefore, we recommend that future research replicate these findings.

Intensity

There was no significant difference between the high-intensity (gun-present) and low-intensity (gun-absent) groups. This may be because the crime's severity did not differ enough as the weapon was not used in the high-intensity condition. As the difference was insignificant, we combined both intensity groups, which acted as a pseudo-replication.

Intercept

Angry voices were generally distrusted, while fearful, neutral, and sad voices were trusted. Disgusted and happy voices were neither trusted nor distrusted. Given that happiness would be inappropriate in this context, it is unsurprising that happiness had no effect here. Prior work has shown that angry people are trusted less (Stouten & De Cremer, 2010). Neutral and sad would-be appropriate emotions in this context, and people displaying those emotions were trusted more. Participants received very little information about the crime and no evidence to support the speaker's testimony; it is interesting to note that audio recordings influenced trust. Past research has identified a growing concern for jurors accepting eyewitness testimony as fact (Fradella, 2006). Our findings support this research showing that mock jurors tend to trust witnesses' testimony even when no supporting evidence is available. As such, judges need to admit expert testimony on the reliability of eyewitness testimony and factors that might impact the accuracy of such testimony when applicable.

Summary

This research is essential as it is the first experiment to examine the interaction between race, sex, and trust across emotions. Most of the published research in the voice-literature field focused White Male voices or unidentified race male (O'Connor et al., 2014; Montano et al., 2017; Tigue et al., 2011; Mayew et al., 2013). Hopefully, our research and the differences in biases found toward the race and sex of speakers will push researchers to diversify their stimuli. This research supports the idea that testing across different emotions provides a more enriched understanding of the human experience. It also highlights judgment biases that may occur in courtrooms which is a crucial step to ensuring that all individuals have an equal prosecution which should be a goal of the criminal justice system. Although our participants said the same sentence while conveying the same six emotions, they were perceived differently, with black males and lower-pitched speakers trusted the least. Bringing further awareness of how deep our biases run will allow judges and jurors to be more cautious when making decisions that have lifelong effects.

Chapter 3: General Discussion

Our study demonstrated that sex, race, emotion, and vocal characteristics impact perceptions of trust of auditory testimonies. Findings that there is a significant interaction between speaker sex and race supported the idea that heuristic cues are used to foster decision-making when evidence is incomplete (Cho et al., 2015). Raters in our experiment perceived speakers saying brief sentences differently, which support theories that individuals make judgments about others within seconds of contact (Bergmann et al., 2012). This is concerning as eyewitnesses are one of the most compelling sources of information judges and jurors receive (Beety, 2022; Fedella, 2006). Considering that we are more likely to trust those who are most like us (Maeder & Yamamoto, 2019), biases and stereotypes disproportionately affect marginalized eyewitnesses, as judges and jury members are typically predominantly White (Johnson, 1985). With Black, Asian, and Native American individuals making up nearly 50% of inmates in the US prison system (Figure 3.1), steps need to be taken to ensure that there is more diversity on juries; especially considering that biases towards heuristic cues do not help individuals in accurately identifying individuals who are trustworthy or untrustworthy (Campellone & Kring, 2013; Schild et al., 2020) or guilty from innocent. Increasing the diversity of judges and jurors may mitigate the effect of stereotypes and mental shortcuts (Maeder & Yamamoto, 2019) on decision-making in the courtroom.

Race	# of Inmates	% of Inmates
Asian	2,280	1.4%
Black	60,936	38.5%
Native American	4,158	2.6%
White	91,095	57.5%

Figure 3.1: Inmate Race data obtained from the Federal Bureau of Prisons. Last updated on Saturday, 18 March 2023 (https://www.bop.gov/about/statistics/statistics_inmate_race.jsp)

The sex and race of speakers significantly impacted perceptions of trust in the courtroom across emotions. Findings of attorneys being perceived differently because of their sex (Hodgson & Pryor, 1984; Salerno et al., 2018) and presented emotion replicated in eyewitnesses. Face perception research pointed to emotions such as happiness increasing perceptions of trustworthiness and emotions such as anger decreasing perceptions of trust (Campellone & Kring, 2013). However, in this experiment angry testimonies were less likely to be trusted in both male and female speakers than other emotions indicating that context may play a prominent role in moderating trust perceptions. Considering that heuristic cues impact legal sentencing decisions (Jaegar et al., 2020; Wilson & Rule, 2016), judges and jurors should know how biases toward speaker sex and emotion impact their decision-making. Individuals make inferences from non-verbal cues such as sex, race, and voice pitch.

Emotion effects the interaction between voice pitch and perceptions of trust. Previous studies have not looked at the impact of emotion on trust perception. While previous studies have found either a preference for lower or higher-pitched voices for males and females (O'Conner et al., 2011; O'Connor et al., 2014; Feinberg et al., 2005; Apicella et al., 2007; Pisanski et al., 2018; Montano et al., 2017; Mayew et al., 2013; Klofstad et al., 2012; Klofstad et al., 2015), preferences for voice pitch varied significantly across emotion and speaker sex. Future studies should explore how emotion moderates trust perception in different contexts.

This experiment adds to the limited research on trust perception and voice characteristics such as CPP, VTL, SHR, Speech rate, LTAS, and HNR. Much like voice pitch, perceptions of trust across additional voice characteristics varied significantly across speaker sex and emotion. As this is the first experiment to conduct an in-depth analysis of trust across all of these voice characteristics, further studies should be added to this body of literature.

This study did not include information on the defendant's race. However, future studies should include this information as researchers have found that in specific racial subgroups, individuals are more likely to be skeptical of eyewitness testimonies because of the interaction between the defendant, juror, and eyewitness race (Maeder & Ewanation, 2018). Future studies should also strive to obtain a more diverse sample group. While we attempted to do this, we did not have enough participants in each racial subgroup to analyze the interaction between speaker race and participant race on perceptions of trust.

Non-verbal vocal properties of sex, race, and emotion impact trust perceptions of eyewitnesses. Uncovering such biases in a courtroom setting is crucial to ensuring that all individuals receive equitable treatment in the criminal justice system. This research is essential as it is the first experiment to examine the interaction between race, sex, vocal characteristics, and trust across emotions. This research also supports the idea that testing across different emotions provides a more enriched understanding of the human experience. It also highlights judgment biases that may occur in courtrooms which is a crucial step to ensuring that all individuals have an equal prosecution which should be a goal of the criminal justice system.

References

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology letters*, 3(6), 682-684.

Bagby, R. M., Parker, J. D., Rector, N. A., & Kalemba, V. (1994). Racial prejudice in the Canadian legal system: Juror decisions in a simulated rape trial. *Law and Human Behavior*, *18*(3), 339.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, 4, 328.

Beety, V. (2022). Considering" Machine Testimony": The Impact of Facial Recognition Software on Eyewitness Identifications. Duq. L. Rev., 60, 271.

Belin, P., Boehme, B., & McAleer, P. (2017). The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PloS one*, 12(10), e0185651.

Bergmann, K., Eyssel, F., & Kopp, S. (2012). A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12* (pp. 126-138). Springer Berlin Heidelberg.

Bucolo, D. O., & Cohn, E. S. (2010). Playing the race card: Making race salient in defence opening and closing statements. *Legal and Criminological Psychology*, *15*(2), 293-303.

Burgoon, J. K., & Le Poire, B. A. (1999). Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communications Monographs*, 66(2), 105-124.

Campellone, T. R., & Kring, A. M. (2013). Who do you trust? The impact of facial emotion and behaviour on decision making. *Cognition & emotion*, *27*(4), 603-620.

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). Crema-d: Crowd-sourced emotional multimodal actors dataset. *IEEE transactions on affective computing*, *5*(4), 377-390.

Chan, M., & Cardenas, J. (2021). We See You: Increasing Inclusion and Empowerment of Immigrants by Liberating Library Workflows from Institutional Biases Rooted in Colonial Frameworks. *Library Juice Press.*

Cho, J. H., Chan, K., & Adali, S. (2015). A survey on trust modeling. ACM Computing Surveys (CSUR), 48(2), 1-40.

Cohn, E. S., Bucolo, D., Pride, M., & Sommers, S. R. (2009). Reducing White Juror Bias: The Role of Race Salience and Racial Attitudes 1. *Journal of Applied Social Psychology*, *39*(8), 1953-1973.

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289-311.

Devine, D. J., & Caughlin, D. E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law,* 20(2), 109.

Edelenbos, J., & Klijn, E. H. (2007). Trust in complex decision-making networks: A theoretical and empirical exploration. *Administration & society*, 39(1), 25-50.

Evelyn M. Maeder & Joel Burdett (2013) The Combined Effect of Defendant Race and Alleged Gang Affiliation on Mock Juror Decision-Making, Psychiatry, Psychology and Law, 20:2, 188-201, DOI: 10.1080/13218719.2011.633330

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental science*, 18(2), 327-334.

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.

Feinberg, D. R., & Cook, O. (2020, September 15). VoiceLab: Automated Reproducible Acoustic Analysis. https://doi.org/10.31234/osf.io/v5uxf

Ferrand, C. T. (2002). Harmonics-to-noise ratio: an index of vocal aging. *Journal of voice*, 16(4), 480-487.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., & King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behavioral Sciences & the Law*, *24*(2), 179-198.

Fradella, H. F. (2007). Why judges should admit expert testimony on the unreliability of eyewitness testimony. Fed. Cts. L. Rev., 2, 1.

Fraile, R., & Godino-Llorente, J. I. (2014). Cepstral peak prominence: A comprehensive analysis. Biomedical Signal Processing and Control, 14, 42-54.

Galinsky, D. F., Erol, E., Atanasova, K., Bohus, M., Krause-Utz, A., & Lis, S. (2020). Do I trust you when you smile? Effects of sex and emotional expression on facial trustworthiness appraisal. *PLoS One*, *15*(12), e0243230.

Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., ... & Sataloff, R. T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology*, *112*(4), 324-333.

Hodgson, S., & Pryor, B. (1984). Sex discrimination in the courtroom: Attorney's gender and credibility. *Psychological Reports*, 55(2), 483-486.

How many innocent people are in prison? Innocence Project. (2023, February 23). Retrieved February 28, 2023, from https://innocenceproject.org/how-many-innocent-people-are-in-prison/

Imhof, M., Välikoski, T. R., Laukkanen, A. M., & Orlob, K. (2014). Cognition and interpersonal communication: The effect of voice quality on information processing and person perception. *Studies in Communication Sciences*, 14(1), 37-44. .

Jaeger, B., Oud, B., Williams, T., Krumhuber, E., Fehr, E., & Engelmann, J. B. (2020). Trustworthiness detection from faces: Does reliance on facial impressions pay off. *Advance online publication*. DOI: https://doi. org/10.31234/osf. io/ayqeh.

Johnson, S. L. (1985). Black innocence and the white jury. *Michigan Law Review,* 1611-1708.

Jones, C. S., & Kaplan, M. F. (2003). The effects of racially stereotypical crimes on juror decision-making and information-processing strategies. *Basic and Applied Social Psychology*, *25*(1), 1-13.

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698-2704.

Klofstad, C. A. (2016). Candidate voice pitch influences election outcomes. *Political psychology*, *37*(5), 725-738.

Knowles, K. K., & Little, A. C. (2016). Vocal fundamental and formant frequencies affect perceptions of speaker cooperativeness. *Quarterly Journal of Experimental Psychology*, 69(9), 1657-1675.

Lang, C. (2020, July 6). *How the Karen Meme confronts history of white womanhood*. Time. Retrieved March 3, 2023, from https://time.com/5857023/karen-meme-history-meaning/

Leong, K., Hawkshaw, M. J., Dentchev, D., Gupta, R., Lurie, D., & Sataloff, R. T. (2013). Reliability of objective voice measures of normal speaking voices. *Journal of Voice*, 27(2), 170-176.

Maeder, E. M., Yamamoto, S., & Saliba, P. (2015). The influence of defendant race and victim physical attractiveness on juror decision-making in a sexual assault trial. *Psychology, Crime & Law*, *21*(1), 62-79.

Maeder, E. M., & Yamamoto, S. (2019). Investigating race salience, defendant race, and victim race effects on mock juror decision-making in Canada. *Justice Quarterly*, *36*(5), 929-953.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion.

Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior*, *34*(4), 243-248.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PloS one*, 9(3), e90779.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of personality and social psychology*, 34(4), 615.

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: a meta-analytic review of defendant treatment. *Law and human behavior*, 29(6), 621.

Montano, K. J., Tigue, C. C., Isenstein, S. G., Barclay, P., & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, *38*(3), 293-297.

Moore, B. J. (1995). Hearing (2nd ed.). San Diego, CA: Academic Press.

Nakano, T., & Yamamoto, T. (2022). You trust a face like yours. *Humanities and Social Sciences Communications*, 9(1), 1-6.

O'Connor, J. J., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology*, *9*(1), 147470491100900109.

O'Connor, J. J., Pisanski, K., Tigue, C. C., Fraccaro, P. J., & Feinberg, D. R. (2014). Perceptions of infidelity risk predict women's preferences for low male voice pitch in short-term over long-term relationship contexts. *Personality and Individual differences*, *56*, 73-77.

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, *40*(1), 1-18.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. 10.3758/s13428-018-01193-y

Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., & Reby, D. (2018). Voice pitch modulation in human mate choice. *Proceedings of the Royal Society* B, 285(1893), 20181634.

Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *The Journal of the Acoustical Society of America*, *101*(4), 2267-2277.

Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and human behavior*, 27(4), 283-296.

Re, D. E., O'Connor, J. J., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PloS one,* 7(3), e32719.

Rhodes, G., Halberstadt, J., & Brajkovich, G. (2001). Generalization of mere exposure effects to averaged composite faces. *Social Cognition*, 19(1), 57-70.

Salerno, J. M., Phalen, H. J., Reyes, R. N., & Schweitzer, N. J. (2018). Closing with emotion: The differential impact of male versus female attorneys expressing anger in court. *Law and Human Behavior*, 42(4), 385.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408-2412.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, *32*(1), 76-92.

Schild, C., Stern, J., & Zettler, I. (2020). Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology*, 31(1), 164-175.

Schild, C., Stern, J., Penke, L., & Zettler, I. (2021). Voice pitch–A valid indicator of one's unfaithfulness in committed relationships?. *Adaptive Human Behavior and Physiology*, 7, 245-260.

Schirmer, A., Feng, Y., Sen, A., & Penney, T. B. (2019). Angry, old, male–and trustworthy? How expressive and person voice characteristics shape listener trust. *PloS one*, 14(1), e0210555.

Siegrist, M., Gutscher, H., & Earle, T. C. (2005). Perception of risk: the influence of general trust, and general confidence. *Journal of risk research*, 8(2), 145-156.

Smith, S. S. (2010). Race and trust. *Annual Review of Sociology*, *36*, 453-475.

Sun, X. (2002, May). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In 2002 *IEEE international conference on acoustics, speech, and signal processing* (Vol. 1, pp. I-333). IEEE. And affects trust judgements

Simpson, J. A. (2007). Psychological foundations of trust. *Current directions in psychological science*, 16(5), 264-268.

Stouten, J., & De Cremer, D. (2010). "Seeing is believing": The effects of facial expressions of emotion and verbal communication in social dilemmas. *Journal of Behavioral Decision Making*, 23(3), 271-287.

Szakay, A. (2012). Voice quality as a marker of ethnicity in New Zealand: from acoustics to perception 1. *Journal of Sociolinguistics*, 16(3), 382-397.

Tortosa, M. I., Strizhko, T., Capizzi, M., & Ruz, M. (2013). Interpersonal effects of emotion in a multi-round Trust Game. Psicologica: *International Journal of Methodology and Experimental Psychology*, 34(2), 179-198.

Tigue, C. C., Borak, D. J., O'Connor, J. J., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210-216.

Velner, E., Truong, K. P., & Evers, V. (2021). Speaking of Trust--Speech as a Measure of Trust. *arXiv* preprint arXiv:2104.05340.

VoiceLab: Software for Fully Reproducible Automated Voice Analysis D. Feinberg Proc. Interspeech 2022, 351-355.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. Social Psychological and Personality Science, 7(4), 331-338.

Wirfs-Brock, A., & Eich, B. (2020). JavaScript: the first 20 years. *Proceedings of the ACM on Programming Languages*, *4*(HOPL), 1-189.

Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. PloS one, 8(4), e62397.