Audio-visual integration in social evaluation

Mila Mileva[1], James Tompkinson[2], Dominic Watt[2], & A. Mike Burton[1]

[1] Department of Psychology, University of York, UK
[2] Department of Language and Linguistic Science, University of York, UK

Correspondence to:
A. Mike Burton
Department of Psychology
University of York
York
YO10 5DD, UK
mike.burton@york.ac.uk

Running head: Audio-visual integration in social evaluation

Word Count: 7,169

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262 to A. Mike Burton.

Abstract

Our social evaluation of other people is influenced by their faces and their voices. However, rather little is known about how these channels combine in forming 'first impressions'. Over five experiments we investigate the relative contributions of facial and vocal information for social judgements: dominance and trustworthiness. The experiments manipulate each of these sources of information *within-person*, combining faces and voices giving rise to different social attributions. We report that vocal pitch is a reliable source of information for judgements of dominance (Experiment 1) but not trustworthiness (Experiment 4). Faces and voices make reliable, but independent contributions to social evaluation. However, voices have the larger influence in judgements of dominance (Experiment 2), whereas faces have the larger influence in judgements of trustworthiness (Experiment 5). The independent contribution of the two sources appears to be mandatory, as instructions to ignore one channel do not eliminate its influence (Experiment 4). Our results show that information contained in both the face and the voice contributes to first impression formation. This combination is, to some degree, outside conscious control, and the weighting of channel contribution varies according the trait being perceived.

Keywords: First impressions; social evaluation; audio-visual integration; faces; voices

Public Significance Statement

This study shows how our first impressions of someone are formed on the basis of their face and their voice. We combine these sources of information automatically, but some judgements are influenced more by faces, and some by voices.

Introduction

A wealth of biological and social information about people, such as sex, age, ethnicity or emotional state, can be inferred by either looking at their faces or listening to their voices (Bruce & Young, 1986; Belin, et al., 2011; Yovel & Belin, 2013). Moreover, we constantly recognise people's identities from their faces and voices, for example by looking at a photograph or hearing a voice on the telephone. People infer socially-relevant information and form stable first impressions about unfamiliar others from both faces and voices (Todorov, Pakrashi, & Oosterhof, 2009; Zuckerman & Driver, 1989). Social impressions from faces arise very quickly (after less than a second of exposure in many reports), whereas impressions from voices will always include some temporal element.

While first impressions might not represent reality accurately, social evaluation is characterised by a high level of agreement between observers or listeners for both facial and vocal information (Zebrowitz & Montepare, 2008; McAleer, Todorov & Belin, 2014), which implies that people use consistent physical information in the face and acoustic information in the voice to inform their social judgements. Moreover, these zero-acquaintance impressions have been repeatedly shown to influence our social interactions and decisions. Voting behaviour, for example, can be influenced by both facial and vocal information, with studies demonstrating that voting outcomes can be predicted by the perceived competence in a candidate's face (Ballew & Todorov, 2007; Olivola & Todorov, 2010) or the pitch of their voice (Tigue, et al., 2012). Similarly, both facial and vocal information have been shown to predict courtroom outcomes (Chen, Halberstam, & Alan, 2016; Wilson & Rule, 2016) as well as to influence dating and mate preferences (Little, Burt, & Perrett, 2006; Wells, et al., 2009).

First impressions from both faces and voices have been shown to fall along two fundamental dimensions, one representing valence and the other representing dominance. In face evaluation, Oosterhof and Todorov (2008) used Principal Component Analysis (PCA) on spontaneous, unconstrained personality descriptors inferred from faces and showed that first impressions can be reduced to trustworthiness and dominance. Likewise, following a similar procedure, McAleer, Todorov and Belin (2014) also demonstrated a two-dimensional space for social evaluation of voices with valence and dominance as the main dimensions. Such findings are consistent with general social evaluation models such as concept evaluation (Osgood, Suci & Tannenbaum, 1957), group evaluation (Fiske, Cuddy & Glick, 2007) and

models of interpersonal perception (Wiggins, 1979), all of which rely on two orthogonal dimensions - affiliation and dominance.

Audio-visual integration

In this study we are concerned with first impressions gained from multimodal stimuli, comprising faces and voices. Given that both these sources individually have been shown to give rise to consistent social attributions, how do they interact? Do voices or faces dominate in social judgements, or does the signal from one source influence the interpretation of the other? In perception of different types of signals, researchers have shown very strong integrative effects. For example, facilitative multimodal influences have been demonstrated in speech intelligibility or 'lip-reading', where presenting participants with visual information from a speaker's face can significantly improve speech content recognition (from 23% to 65% in Summerfield, 1979). In person identification studies, participants are quicker to identify a face as familiar after being presented with the voice of that same identity and vice versa (Ellis, Jones & Mosdell, 1997; Schweinberger, Herholz & Stief, 1997). A classic interference effect comes from the McGurk illusion (McGurk & MacDonald, 1976) in speech perception, whereby participants are presented with incongruent audio and visual cues and yet integrate them together. Attending to a video clip of a person pronouncing the syllable /ga/ while listening to a superimposed audio clip of a person pronouncing the syllable /ba/, for example, commonly results in the impression that the person in the video clip actually pronounces the syllable /da/.

While both voices and faces provide us with a wealth of social information (Bruce & Young, 1986; Belin, et al., 2011) and there is a multitude of studies investigating the independent effects of facial and vocal cues on social perception (Oosterhof & Todorov, 2008; Hodges-Simeon, Gaulin & Puts, 2010; Berry, 1990; Zuckermann & Driver, 1989), existing audio-visual integration research has been almost exclusively focused on emotion and identity recognition (see Campanella & Belin, 2007 for a review). Massaro and Egen (1996), for example, presented participants with congruent and incongruent face-voice pairings where face images displayed happy, angry or neutral expressions, while the voice stimuli were created by an actor pronouncing the word *"please"* in a happy, angry or neutral way. Participants' task was simply to classify the emotion as happy or angry. The study showed

that while both facial and vocal cues were effective for expression categorisation, visual information from the face had a stronger effect, as it changed participants' performance across all three voice emotion levels, results which are consistent with the general finding that faces seem to be more reliable cues than voices in emotion recognition (Mehrabian & Ferris, 1967; Hess, Kappas & Scherer, 1988).

There is evidence that audio-visual integration in emotion recognition is an automatic process, as participants seem to incorporate face and voice cues together even when they are instructed to ignore one of the information channels. For example, de Gelder and Vroomen (2000) found a significant effect for both the visual and vocal channels on the perception of happiness/sadness and happiness/fear when participants were presented with both channels but specifically instructed to ignore either the face or the voice when making their judgements. Evidence for the automatic nature of audio-visual integration also comes from studies on identity recognition (Campanella & Belin, 2007). In a series of experiments Schweinberger, et al. (2007, 2011) demonstrate that presenting participants with corresponding and non-corresponding face-voice pairs had an influence on familiarity decisions: recognition of a familiar voice was faster and more accurate when it was paired with the corresponding face - even when participants were specifically instructed to base their decisions exclusively on the audio cues.

In comparison with research examining emotion and identity recognition from faces and voices, comparatively fewer studies have explored the effect of combining visual and vocal cues on the formation of first impressions. This is in spite of features such as dominance, trustworthiness and attractiveness forming a key part of prominent social perception models (Fiske et al, 2007; Oosterhof and Todorov, 2008). Rezlescu et al (2015) examined listener perceptions of attractiveness, trustworthiness and dominance using a combination of static male faces and brief vowel sounds produced by male speakers adopting a variety of emotional vocal expressions such as happy, sad and angry. The results indicated that facial information was more influential in judgements of attractiveness, whereas vocal information was more influential in dominance judgements. Both visual and vocal information contributed significantly to trustworthiness judgements. However, Tsankova, et al. (2015) examined perceptions of trustworthiness using facial and vocal cues and argued that trustworthiness judgements were more heavily influenced by facial information than by vocal information.

Research aims

Comparatively little is known about the combined effects of vocal and facial cues on social evaluation. In the studies below we aim to investigate the relative contribution of audio and visual information to the perception of the fundamental social perception dimensions - trustworthiness and dominance. We also aim to explore whether this audio-visual integration is automatic.

Our approach differs from that taken in previous studies in that we use vocal stimuli comprising speech, which (arguably) represent real-world social interactions more accurately than non-verbal vocalisations. While some argue that the use of brief, neutral vowel sounds mitigates the influence of aspects of voice such as prosody and semantic content (Rezlescu, 2015), the extent to which using non-verbal vocalisations replicates real everyday speech has been the topic of debate (Apple et al, 1979). Social evaluations are clearly multi-faceted in everyday life, and so there is value in studying them using contentful utterances.

Our approach also differs from previous work in that we make use of within-person variability to manipulate social person evaluations. In most studies of first impressions, it is assumed that *people* give rise to stable judgements, i.e. a particular person is judged more or less trustworthy, dominant etc. However, this is now known to be false. Ratings for different photos of the same person can vary more than for photos of different people (Jenkins et al, 2011; Todorov & Porter, 2014). First impressions derived from faces can therefore reflect differences in *photos* rather than differences in *people* (Burton, 2013; Burton et al, 2016). Rather than using different identities rated as high or low in dominance and trustworthiness, here we sample different images of the same identity and select those rated as the most and least trustworthy and dominant.

We also isolate the effect of a single acoustic measure – mean pitch - which has previously been linked to perceptions of dominance and trustworthiness in voices (Ohala, 1984; Tsanani, 2016). In Study 1 we first validate a set of vocal stimuli and investigate the role of pitch in dominance perception. In Study 2, these auditory stimuli were matched with a set of face images perceived as high and low in dominance to investigate the relative effects of both channels on social person perception. Study 3 extends work on the automaticity of audio-

visual integration (de Gelder & Vroomen, 2000; Schweinberger, 2007) into the domain of first impressions. We present participants with both facial and vocal cues and instruct them to ignore one of those channels when they evaluate each person. Studies 1-3 focus on perceptions of dominance. In Studies 4 and 5 we extend these into perception of trustworthiness. In Study 4 we evaluate the use of pitch as a cue to trustworthiness, and in Study 5 we examine multimodal trustworthiness perception.

Study 1: Perception of dominance from voices

Overview

This first experiment was conducted to obtain baseline judgements of dominance for auditory stimuli, independent of visual information. The specific vocal parameter investigated in this study is mean fundamental frequency (F0), which we label *mean pitch*. Although the link between pitch and F0 is strictly non-linear, Laver (1994) argues that at the low frequencies relevant for the perception of pitch in both male and female voices, a linear relationship can be assumed. We manipulated the pitch of vocal stimuli, hypothesising that this would affect perception of dominance. Pitch has been highlighted as one of the most perceptually salient acoustic cues used by listeners to infer emotion and affect in speech (Dimos, et al., 2015). Following work which identifies low pitch as a signal of aggression and dominance across a variety of animal species (Morton, 1977), research has identified a link between the lowering of F0 and the perception of both social and physical dominance in human speech (Ohala, 1984; Puts et al, 2006; Puts et al, 2007; Tusing and Dillard, 2000).

It is important to establish whether pitch manipulation has the hypothesised effect in the perception of verbal stimuli produced by male and female speakers. The previous literature is somewhat contradictory, perhaps reflecting the wide diversity in the types of stimuli used (Borkowska & Pawlowski, 2011; McAleer, et al., 2014; Tsantani, et al., 2016; Vukovic, et al., 2011). To anticipate the results, we found that verbal utterances were judged more dominant when rendered with lower pitch, an effect which held for both male and female voices.

Method

Participants

Voices were rated by 36 participants (13 male, mean age = 23.9, age range = 18-36). Sample size was based on McAleer et al. (2014) in which ratings were gathered from 32 participants per trait. Four extra participants were tested as they signed up for the study before the end of the recruitment period. All participants were students at the University of York and received payment or course credits for their participation. Informed consent was provided prior to participation in accordance with the ethical standards stated in the 1964 Declaration of Helsinki.

Materials

Experimental stimuli were 40 voice recordings (2 for each of 20 identities, one manipulated to a higher pitch and the other manipulated to a lower pitch). Twenty speakers (10 male, mean age = 23, age range = 18-35) gave informed consent to be recorded producing the utterance *"I wouldn't do that if I were you"*. Voices were recorded following ethical consent from the Department of Language and Linguistic Science at the University of York. All speakers were students at the University of York. Recordings were conducted in quiet recording environments using a Zoom H4N handheld recorder, with the built-in microphone positioned 30cm from each speaker.

The utterance *"I wouldn't do that if I were you"* was chosen due to its indirect nature (Searle, 1979) and because it can give rise to a range of social inferences, including interpretations that it represents advice or threat. Our approach therefore differs from those based on presentations of neutrally-worded reading passages or non-verbal vocalisations (e.g. vowels sounds), both of which are very commonly used techniques in this field (Berry, 1991; Rezlescu, et al., 2015).

Digital manipulations using Praat (Boersma and Weenink, 2016) were used in order to create contrasting mean pitch levels for each stimulus. A Praat pitch alteration script (Fecher, 2015) was used to create low and high mean pitch levels. For male speakers, the mean F0 of each recording was altered to 90Hz (low) and 140Hz (high). These values are 25Hz above and below an approximation of the average male mean F0 level (Hudson et al 2007; Künzel, 1989; Lindh, 2006), and represent values in the highest and lowest 10% of population values reported by Hudson et al (2007). For female speakers, the mean F0 of each recording was altered to 170Hz (low) and 250Hz (high). These values are 40Hz above and below the approximation of an average female F0 level, and reflect the low and high ends of the mean

F0 range reported for female speakers (Künzel, 1989; Traunmüller and Erickson, 1995). All recordings were checked to ensure that no digital artefacts had influenced the sound quality as a result of the editing process. The alteration procedure also preserves the shape of the intonation contour and pitch range whilst altering the mean pitch level.

Procedure

Data were collected online using Qualtrics software (2015, Provo, UT). Participants were presented with each recording individually and asked to rate dominance on a scale from 1 (not at all dominant) to 9 (extremely dominant). Participants rated all 40 of the vocal stimuli, each in an independently randomised order.

Results and discussion

Dominance ratings had very high inter-rater reliability (Cronbach's $\alpha = .89$). A paired t-test showed that low-pitched voices ($M = 4.82$, $SD = 1.05$) were perceived as significantly more dominant than high-pitched voices ($M = 3.80$, $SD = 1.09$), $t\,(35) = 6.81$, $p < .001$, $d_{rm} = 1.13$[1]. This is consistent with previous studies investigating the effect of vocal pitch on the perception of dominance and aggression (Ohala, 1984).

Despite an overall effect of pitch on perceived dominance, some work with different types of stimuli has suggested that such effects are modulated by speaker gender (McAleer, et al., 2014; Tsantani, et al., 2016). This was not the case for our stimuli, which showed a consistent effect of pitch manipulation for both male speakers (Means: 4.39 vs 5.31; $t\,(35) = 4.87$, $p < .001$, $d_{rm} = .81$) and female speakers (Means: 3.22 vs 4.34; $t\,(35) = 5.94$, $p < .001$, $d_{rm} = .99$).

Having established that the pitch manipulation has the hypothesised effect – i.e. that it is possible to make the same voice sound more or less dominant – we now progress to multimodal experiments in which we combine faces and voices.

[1] We use d_{rm} (Morris and DeShon, 2002), as this measure of effect size controls for correlations between conditions,

Study 2: Multimodal perception of dominance from faces and voices

Overview

In this study we use the vocal recordings validated in Study 1 and pair them with a set of facial stimuli, in order to explore how face and voice evaluations come together to form an integrated impression of dominance. Rezlescu, et al. (2015) report that when participants were required to make dominance judgements to multimodal stimuli (face-voice), their judgements were more influenced by the voices than the faces (a pattern which was reversed for ratings of attractiveness). Our study therefore builds on this finding, but with the following differences.

First, our manipulations of stimulus dominance are not confounded by identity. So, here we present high and low-dominance versions of *the same voices*, as prepared by the pitch manipulation described in Study 1. We also present high- and low-dominance versions of *the same faces* by picking images which had been independently rated. Second, our study uses voices articulating contentful speech, as described in Study 1. Participants hear the same phrase uttered across all combinations of conditions, rather than hearing the content-free vocalisations of some earlier studies. This has the advantage that the speech signal is meaningful – while avoiding any confounding of condition with content.

To anticipate the results, we found additive effects of face and voice on overall judgements of dominance. Dominance of both faces and voices independently contributed to the impression formed when stimuli were presented multimodally. However, consistent with Rezlescu et al (2015) we found that voices had the larger effect on overall judgements.

Method

Participants

64 participants (16 male, mean age = 21.9, age range = 18-32) took part in the study. Sample size was based on effect sizes from Rezlescu, et al. (2015), who reported main effects of face and voice on dominance ratings of η_p^2 = .17 (f(U) = .45) and η_p^2 = .53 (f(U) = 1.06) respectively, demonstrating rather large effects (Cohen, 1988). Using the lowest effect size as a starting point a power analysis using GPower (Erdfelder, Faul & Buchner, 1996) indicated that a sample of 31 participants would be needed to detect an effect of a similar size, with 95% power using a within-subjects ANOVA and alpha at .05. This sample size was then

doubled due to the counterbalancing of different face/voice pairings with all possible pairings being rated by a total of 32 participants. All participants were students at the University of York. All participants had normal or corrected-to-normal vision, reported no hearing impairments and received payment or course credit for their participation. Informed consent was provided prior to participation and experimental procedures were approved by the ethics committee of the Psychology Department at the University of York.

Design

This study used a 2 (face/voice) x 2 (high/low dominance) design. All participants completed 40 trials (10 per condition) in which a face and a voice were presented together, meaning that over the session, participants saw two different images of each stimulus person's face, and heard two different versions of each stimulus person's voice. Face and voice stimuli were not of the same identities, however they were matched for age and gender. Across the experiment, trials were counterbalanced such that all combinations of high-/low-rated faces and voices were presented equally often. Trial presentation order was randomised independently for each participant.

Materials

Voice recordings from Study 1 were used as audio stimuli. Face stimuli were selected from a database of 400 images comprising 20 images each of 20 unfamiliar identities downloaded from an internet search. Images were highly variable or 'ambient' (Jenkins et al., 2011) and therefore captured a great amount of variability within each identity due to different lighting conditions, emotional expressions, pose, etc. (see Figure 1 for examples). Twenty participants (different people from those in the main part of the experiment) rated all 400 of these images for trustworthiness and dominance on a scale from 1 (not at all dominant/trustworthy) to 9 (extremely dominant/trustworthy). Consistent with previous studies, there was high inter-rater reliability (Cronbach's $\alpha = .86$), confirming that the levels of consensus in ratings is high in this set, as normally reported in the literature.

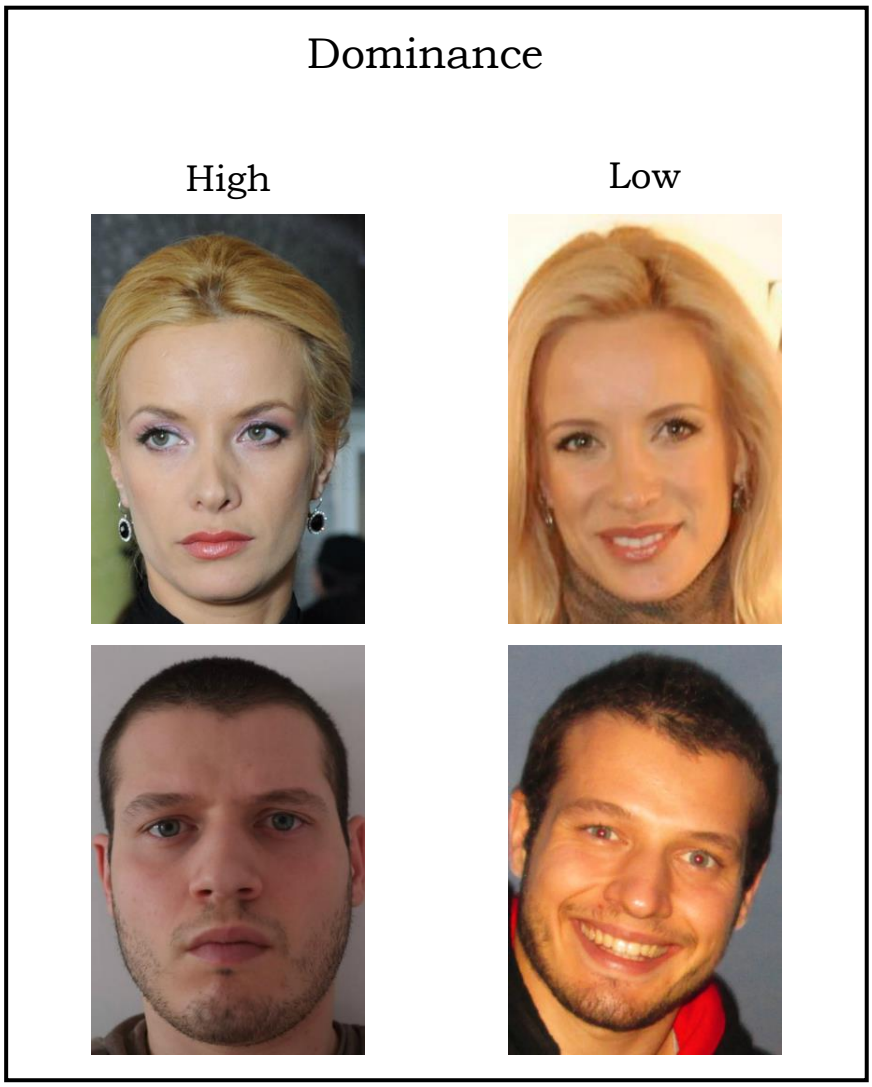

Figure 1. Different images of the same people rated as high and low in dominance

For the purposes of the present study we selected the images that were rated as the most and least dominant for each identity. This yields sets of 20 high- and 20 low-dominance images, with the same identities in each set. Paired t-tests confirmed that images in the high dominance group (M = 6.47, SD = .62) were perceived as significantly more dominant than those in the low-dominance group (M = 4.05, SD = .55, t (19) = 17.48, $p < .001$, d_{rm} = 3.94).

Procedure

Each trial comprised a face and a voice presented simultaneously. The vocal stimuli were played automatically through closed-cup headphones and were presented once only. Participants' task was to rate each identity for dominance on a scale from 1 (not at all dominant) to 9 (extremely dominant). Face stimuli were presented on a white background at

the centre of the screen and the rating scale was positioned below the face image. Participants indicated their response by pressing the corresponding key on the keyboard. The task was not timed, and participants were given no further definition of 'dominance', but encouraged to rely on their 'gut feeling' (Oosterhof & Todorov, 2008).

Results and discussion

Mean ratings by condition are shown in Table 1. A 2x2 within-subjects ANOVA revealed significant main effects of face dominance (F (1, 63) = 72.23, $p < .001$, $\eta_p^2 = .53$) and voice dominance (F (1, 63) = 250.92, $p < .001$, $\eta_p^2 = .80$), with no interaction (F (1, 63) < 1, $\eta_p^2 = .01$).

Table 1

Mean ratings of dominance across conditions in Study 2. SDs in parentheses.

	Low dominance voice	High dominance voice
Low-dominance face	4.0 (.46)	5.1 (.58)
High-dominance face	4.6 (.47)	5.8 (.47)

Our results show clear, independent contributions of face and voice on dominance judgements for multimodal stimuli. Interestingly, the two sources of information do not interact, but provide completely additive contributions to the overall judgement. This is consistent with the findings of Rezlescu et al. (2016), who found no correlations between judgements of dominance from the faces and voices of the same people, thus providing compelling evidence against the validity of these attributions, despite their strong consensus (as replicated here). We also show a similar effect of information source to that of Rezlescu et al. (2016). While both face and voice predict overall dominance ratings, the voice manipulation produces a larger effect. This is consistent with earlier findings on the importance of auditory information for the perception of dominance and aggression, and could be explained by its higher reliability. Dominance judgements have been shown to correlate highly with sexually dimorphic aspects, and vocal pitch is a sexually dimorphic

aspect of voice (Puts et al, 2006). It might, therefore, be a more reliable channel when assessing someone's masculinity, which is related to dominance (Collignon, 2008).

Our results suggest a rather straightforward, additive, system of audio-visual integration for perception of dominance. Two questions therefore arise. In the following experiment we ask how automatic this process is, i.e. to what extent can one weigh either source of evidence through top-down control? Following this, we then return to first impressions more generally, and ask whether this same pattern of additive effects exists for other social judgements.

Study 3: How mandatory is the combination of face and voice in social judgement?

Overview

In the study of emotion perception, there is clear evidence that cues from voices and faces are combined to some extent in a mandatory way. For example, when presented with multimodal stimuli (face and voice) and asked to make a judgement about the person's emotional state, participants incorporate both voice and face cues, even when instructed to base their judgements on just one of these sources (de Gelder and Vroomen, 2000). In the current experiment, we ask whether there is similarly a level of automaticity in cue combination when making judgements of dominance – i.e. making a social judgement rather than an emotional one. To do this, we replicate Study 2, but this time instruct participants to base their judgements on just one of the cues, either voices or faces. If they are able to do this, i.e. by ignoring a competing cue from another channel, it will provide evidence against mandatory combination of cues. To anticipate the results, we find evidence in favour of some mandatory cue combination, based on the result that participants' judgements are consistently influenced by the cues they are instructed to ignore.

Method

Participants

80 participants (8 male, mean age = 19.6, age range = 18-32) from the University of York took part in the study. All had normal or corrected-to-normal vision, reported no hearing impairments and received payment or course credit for their participation. Sample sizes were chosen following Experiment 2, in which effects were larger than those reported in Rezlescu, et al. (2015). A post hoc power analysis was conducted using GPower (Erdfelder, Faul &

Buchner, 1996). This revealed that using a sample of 20 participants would be sufficient to detect such large effects with more than adequate power (>.90, alpha at .05). Participants were randomly assigned to the 'focus on the face' or 'focus on the voice' condition and to one of two different stimuli groups within each condition, meaning that each face/voice pairing was rated by 20 participants.

Design and Procedure

The experiment followed exactly the same procedure as Experiment 2, using the same materials. As above, participants were shown 40 multimodal stimulus trials (face and voice), and asked to make a judgement about the person's dominance. However, in this case half the participants were instructed to make their judgements based on the face only, and the other half to make their judgements on the voice only. Participants were allocated to one of the two groups at random, and all other counter-balancing and trial sequence randomisation was the same as Experiment 2.

Results and discussion

Mean ratings by condition are shown in Figure 2. A three-way mixed-design ANOVA (Instructions: focus on face vs voice; high vs low face dominance; high vs low voice dominance) showed significant main effects of face type (F (1, 78) = 185.29, p < .001, η_p^2 = .70) and voice type (F (1, 78) = 193.71, p < .001, η_p^2 = .71), but no significant three-way interaction, (F (1, 78) = 1.22, p > .05, η_p^2 = .02). Although we did not find a significant main effect of instructions (F (1, 78) < 1, p > .05, η_p^2 = .01), two-way interactions between instructions and face type (F (1, 78) = 69.52, p < .001, η_p^2 = .47) and instructions and voice type (F (1, 78) = 83.14, p < .001, η_p^2 = .52) were both significant. Across instruction conditions face type had a much stronger effect when participants were instructed to focus on the face (F (1, 78) = 240.90, p < .001, η_p^2 = .76) than when they were instructed to focus on the voice (F (1, 78) = 13.91, p < .001, η_p^2 = .15). The same pattern was observed for the effect of voice type – it was much stronger when participants were instructed to focus on the voice (F (1, 78) = 265.33, p < .001, η_p^2 = .77) than when they were instructed to focus on the voice (F (1, 78) = 11.52, p < .01, η_p^2 = .13), showing that participants followed the instructions of the experiment. More importantly, the channel that participants were instructed to ignore nevertheless had a significant effect on their dominance ratings, demonstrating that audio-visual integration is an automatic process that can be controlled to some but not to a complete extent.

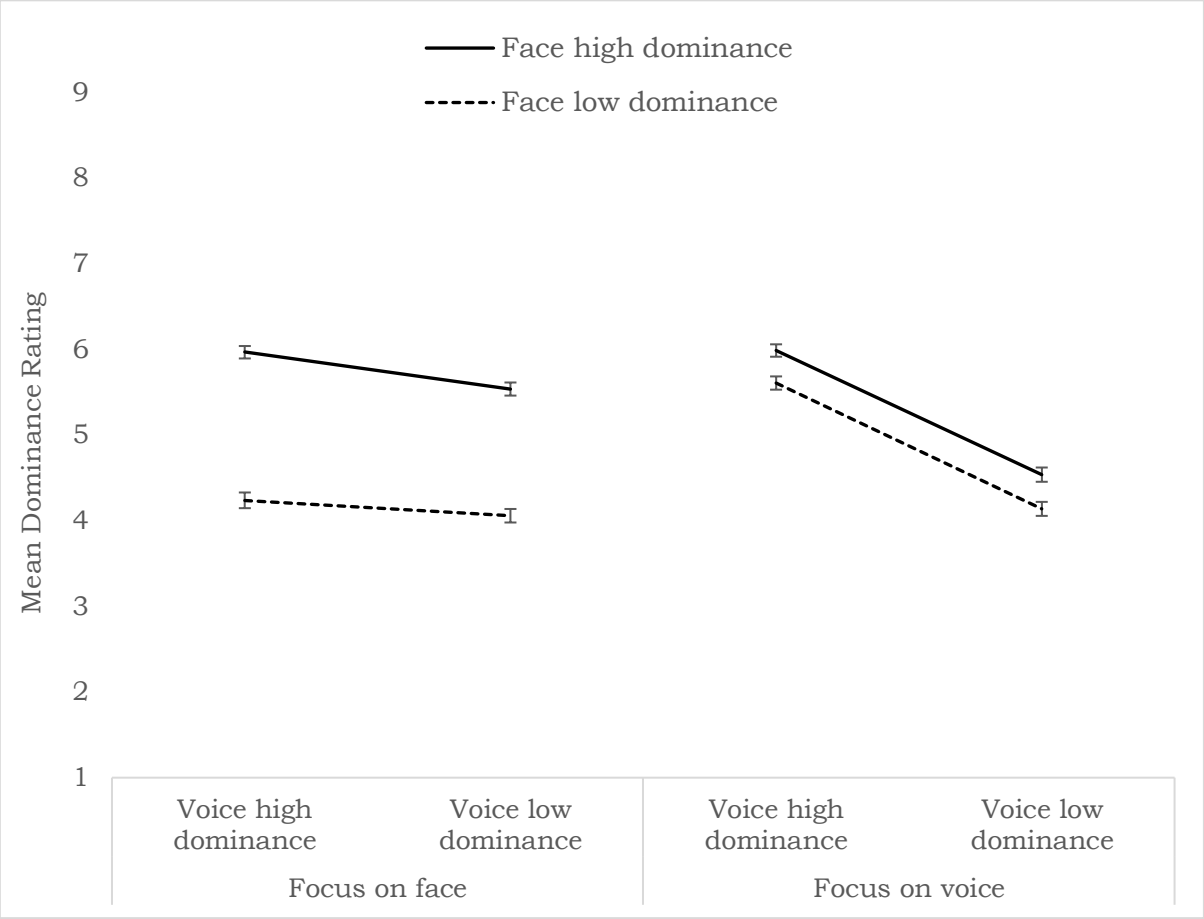

Figure 2. Mean dominance ratings for face-voice pairings under different instructions. Error bars are within-subjects standard error (Cousineau, 2005).

These results show two interesting effects. First, the instructions clearly influenced participants' behaviour. When instructed to focus on faces, the face type had the largest effect on dominance ratings. Similarly, when instructed to focus on voices, the voice type had the largest effect on ratings. Second, and despite this, the cue which participants were instructed to ignore nevertheless had a significant effect on dominance ratings in each case. Furthermore, the effect was independent of the attended cue – there was no significant interaction between attended and ignored cue in either case. These results provide quite clear evidence for some degree of automaticity in the combination of multimodal information in social judgements of dominance. It would appear that the pattern reported in previous work for multimodal perception of emotions (de Gelder & Vroomen, 2000; Schweinberger, et al., 2007) also holds for social impressions.

So far, we have concentrated primarily on the perception of dominance. We have shown that this attribution is made by independent contributions from voices and faces, and there is some degree of mandatory combination of these. In the next two experiments we examine a different social judgement, trustworthiness. We ask whether the pattern of multimodal combination is the same for this judgement as it is for perception of dominance.

Study 4: Perception of trustworthiness from voices

Overview

In study 1, we demonstrated that pitch manipulation affects the perception of dominance in voices when the speaker produces meaningful utterances. In order to study the multimodal perception of trustworthiness (Study 5, below), we first need to establish whether a simple voice manipulation gives rise to reliable changes in perception of this dimension. In fact, there are some reasons to believe that simple pitch manipulation will alter perception of trustworthiness, as it does for dominance. For example, Tsantsani et al. (2016) report a tendency for hearers to judge lower-pitched voices as more trustworthy, both in male and female voices, albeit for temporally reversed speech. However Vukovic et al. (2011) found no effect of pitch on trustworthiness judgements. Here we examine whether the voice samples used in Study 1 – in which pitch is raised or lowered for a spoken sentence - will also give rise to differences in trustworthiness judgements.

Method

Participants

Voices were rated by 38 participants (10 male, mean age = 21.55, age range = 18-35). As with Experiment 1, sample size was based on McAleer et al (2014), in which 32 participants gave ratings. The additional 6 extra participants signed up for the study before the end of recruitment period. All participants were students at the University of York and received payment or course credits for their participation. Experimental procedures were approved by the ethics committee of the Department of Language and Linguistic Science at the University of York.

Materials and Procedure

Experimental stimuli were the same 40 voice recordings as those used for Study 1, i.e. 2 for each of 20 identities, one manipulated with a higher pitch and the other manipulated with a lower pitch. Once again, data were collected online using Qualtrics software. Participants were presented with each recording individually and were asked to rate it for trustworthiness on a scale from 1 (not at all trustworthy) to 9 (extremely trustworthy). The order of stimuli was randomised independently for each participant.

Results and discussion

Trustworthiness ratings had very high inter-rater reliability (Cronbach's $\alpha = .93$). However, there was no difference between trustworthiness ratings for high- ($M = 5.08$, $SD = .61$) and low- ($M = 5.00$, $SD = .60$) pitched voices ($t\ (19) = 1.07$, $p > .05$, $d_{rm} = .25$), regardless of speaker gender. On this basis, we cannot use manipulated versions of the same voice in order to study multimodal perception of trustworthiness. For this reason, in the final study, below, we selected natural stimulus voices which had been independently rated as being high or low in trustworthiness.

Study 5: Multimodal perception of trustworthiness from faces and voices

Overview

In this final study we replicated the approach taken in Study 2 by presenting participants with face-voice parings, and asking them to judge the trustworthiness of the person depicted. Faces and voices which had previously been rated as high or low in trustworthiness were presented in all combinations (high/low face/voice). To anticipate the results, we found independent effects of face and voice trustworthiness, with ratings being influenced more by faces than voices.

Method

Participants

40 participants (8 male, mean age = 20.1, age range = 18-30) took part in the study. Sample size was determined by the same power analysis used for Experiment 3, demonstrating that a sample of 20 participants per counterbalancing group would be enough to detect the large face and voice effects. All were students at the University of York. All participants had

normal or corrected-to-normal vision, reported no hearing impairments and received payment or course credit for their participation. Informed consent was provided prior to participation and experimental procedures were approved by the ethics committee of the Psychology Department at the University of York.

Design

This study used a 2 (face/voice) x 2 (high/low trustworthiness) design. All participants completed 40 trials (10 per condition) in which a face and a voice were presented together, meaning that over the session, participants saw two different images of each stimulus person's face, and heard two different versions of each stimulus person's voice. Across the experiment, trials were counterbalanced such that all combinations of high-/low-rated faces and voices were presented equally often. Trial presentation order was randomised independently for each participant.

Materials

The voice recordings from Study 4 were used as audio stimuli. We performed a median split on ratings of trustworthiness, separately for male and female voices. Combining male and female voices into high- and low-trustworthy groups gives means of 5.48 and 4.61 respectively (*SDs* = .33 and .47), a highly reliable separation (t (19) = 12.05, $p < .001$, d_{rm} = 2.96). Note, that the results of Study 4 require that identities are no longer unconfounded with the voice stimulus dimension. The high- and low-rated stimulus groups contain some voices of the same people, albeit manipulated to different pitches.

Face stimuli come from the same database as that used in Study 2 (20 images of 20 people), and all images were rated for trustworthiness by the same 20 raters, who did not take part in the main experiments. Once again, we used a 9-point scale, from 1 (not at all trustworthy) to 9 (extremely trustworthy). Inter-rater reliability was very high (Cronbach's $\alpha = .94$). To create high- and low-trustworthy groups, we selected the image for each individual which received the highest and lowest mean ratings. Figure 3 shows examples. Paired t-tests confirmed that images in the high trustworthiness group (M = 6.29, SD = .46) were perceived as significantly more trustworthy than those in the low trustworthiness group (M = 4.58, SD = .46), t (19) = 15.69, $p < .001$, d_{rm} = 3.51.

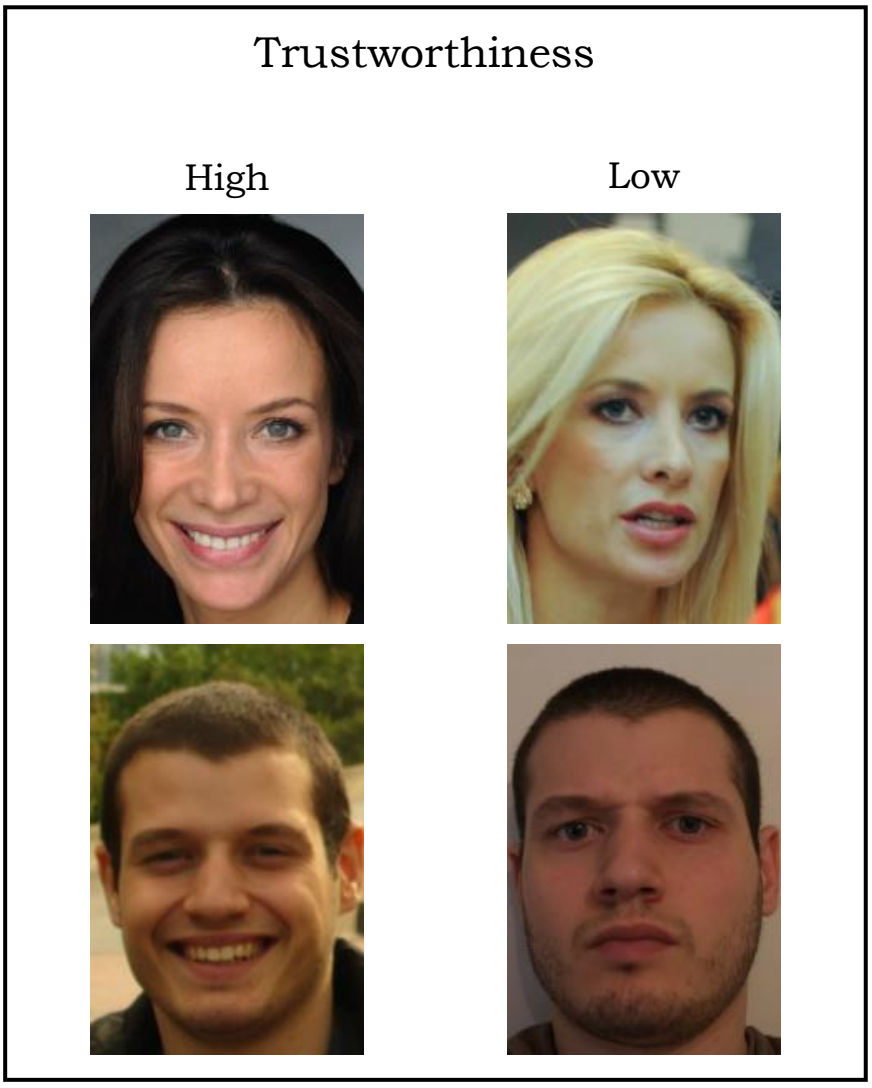

Figure 3. Different images of the same people rated high and low in trustworthiness.

Procedure

Each trial comprised a face and a voice presented simultaneously. The vocal stimuli played automatically and were presented once only. Participants' task was to rate each identity for trustworthiness on a scale from 1 to 9. Face stimuli were presented on a white background at the centre of the screen and the rating scale was positioned below the face image. Participants indicated their responses by pressing the corresponding key on the keyboard.

Results and discussion

Mean ratings by condition are shown in Table 2. A 2x2 within-subjects ANOVA revealed significant main effects of face trustworthiness (F (1, 39) = 99.64, $p < .001$, $\eta_p^2 = .72$) and

voice trustworthiness (F (1, 39) = 18.03, $p < .001$, η_p^2 =.32), with no significant interaction (F (1, 39) = 3.19, $p > .05$, η_p^2 = .08).

Table 2

Mean ratings of dominance across conditions in Study 5. SDs in parentheses.

	Low trustworthiness voice	High trustworthiness voice
Low trustworthiness face	4.8 (.55)	5.4 (.51)
High trustworthiness face	6.0 (.47)	6.2 (.53)

As with judgements of dominance (Study 2), we here show clear, independent contributions of face and voice to multimodal judgements of trustworthiness. However, unlike judgements of dominance, we see in this study that faces have the larger effect when attributions of trustworthiness were being made. This is consistent with findings from correlational studies which show that the judgement of multimodal stimuli can be influenced more or less by faces and voices, according to the attribute involved (Rezlescu et al, 2015).

General Discussion

In a series of experiments we investigate the effect and automaticity of audio-visual integration in social trait attribution. Our results demonstrate that mean vocal pitch is a significant factor in the perception of dominance in voices and that large within-person differences exist in social attribute ratings for faces. Moreover, while both face and voice cues influenced social trait attribution significantly, the relative contributions of the auditory and visual channel to social evaluation were shown to be dependent on the specific social trait. While vocal information was more diagnostic for dominance perception, face information was more diagnostic for the perception of trustworthiness. We also show that audio-visual integration is, to some extent, automatic and that participants cannot completely ignore either the audio or visual channel, even when they are instructed to do so.

Results from the present studies reflect findings from previous research which highlights a stronger and more consistent link between mean pitch and dominance perception than

between pitch and trustworthiness perception. Our findings further extend the literature by demonstrating that lowered pitch is associated with perceptions of higher dominance, regardless of the gender of speaker. This is consistent with Ohala (1982) as well as some more recent studies (Borkowska & Pawlowski, 2011; Jones, et. al, 2010; however, see McAleer et. al, 2014 for different findings). Therefore, not only is pitch an important signal in determining the age, gender or mood of a speaker (Latinus & Belin, 2011), it seems that it can also influence the perception of key social attributes such as dominance. Research on pitch and trustworthiness perception is much less consistent, with some studies reporting that lower pitch leads to higher ratings of trustworthiness (Tigue, 2012), some reporting higher pitch to be perceived as more trustworthy (McAleer, 2014) and others failing to find any association between pitch and trustworthiness (Klofstad, 2012; Vukovic, et. al, 2011). Our findings are consistent with the latter group of studies as we did not find a significant association between pitch and trustworthiness. Nevertheless, pitch is one of many acoustic vocal parameters and our audio-visual integration studies show that vocal information has a significant effect on trustworthiness attribution. This implies there might be other acoustic measures worth exploring, such as harmonic-to-noise ratio, which has previously been found to predict ratings of trustworthiness for both male and female speakers (McAleer, 2014).

In terms of multimodal social evaluation, our results show clear differences in the relative contributions of auditory and visual cues to social perception for the two fundamental social dimensions, trustworthiness and dominance. Both the face and the voice had a significant effect on trait attribution. However, while audio information was much more diagnostic of dominance perception, the reverse pattern was observed for trustworthiness, for which face cues were much more important. Our results for multimodal dominance perception replicate and support the findings of Rezlescu et al. (2016). However, they show an interestingly different pattern of results to reports of the facial overshadowing effect (Tomlin, Stevenage, & Hammond, 2016), an advantage for visual information in *identity* recognition. This highlights the importance of both context and task demands, and is consistent with face and voice models proposing that identity, affect and speech information is processed along functional pathways which are mostly independent, yet have some scope to interact with one another (Belin, et. al, 2011; Young & Bruce, 2011). The importance of auditory information to the perception of dominance and aggression could be due to its higher reliability. Dominance judgements have been shown to correlate highly with sexually dimorphic aspects of human physical attributes and behaviour, and vocal pitch is a sexually dimorphic aspect of

voice (Puts et al, 2006). It might, therefore, be a more reliable channel when, for example assessing someone's masculinity, which is related to dominance (Collignon, 2008).

Our findings regarding trustworthiness perception, on the other hand, are in contrast to those of Rezlescu et al (2016), who found that the facial and vocal channels contributed equally to the perception of trustworthiness and interacted with one another. This might be due to the different facial and vocal stimuli used in the present study, as we opted to use contentful speech rather than brief neutral vowel sounds. A consistent finding in the face evaluation literature is that social judgements are highly dependent on emotional expressions and that participants often assign a particular emotional expression to seemingly neutral faces (Said, Sebe, & Todorov, 2009). Our findings may therefore indicate that the visual channel is more reliable than the vocal channel for extracting emotional content (Massaro & Egen, 1996).

We have also shown that the combination of auditory and visual cues is mandatory and bidirectional. Such results are consistent with studies of audio-visual integration in emotion and identity recognition (de Gelder & Vroomen, 2000; Schweinberger, et al., 2007), all of which imply that combining cross-modal information is not under attentional control. It would appear that presenting faces and voices together, regardless of task and synchronicity, leads to an automatic integration rather than prompting perceivers to make a decision about whether or not to integrate the presented information. The evidence for the automaticity of audio-visual integration is particularly compelling here, as the voices in the present studies were paired with static faces. While this method unquestionably misrepresents real-life social interactions, it provides a clear indication of the magnitude of the automaticity effect – a finding further supported by studies reporting automatic integration even when there was a mismatch in the gender of the face and the voice that participants were presented with (Green, Kuhl, & Meltzoff, 1991).

In conclusion, our findings demonstrate clear differences in the weighting of auditory and visual cues in social perception, dependent on the specific social attribute being evaluated. While vocal information is more important for the perception of dominance, facial information has a greater influence on listener attributions of trustworthiness. Furthermore, using a focused-attention paradigm, we show that audio-visual integration appears to be an automatic, bidirectional process. Such findings extend and contribute to the scarce literature on multimodal social person evaluation. By using contentful utterances as vocal stimuli, we

obtained listener evaluations of speech that represent everyday social interactions more accurately. Moreover, we used images of the same people in both the high- and low-dominance and trustworthiness conditions and found significant differences between them. This demonstrates that sufficient within-person variability exists in ratings of different images of the same identity, and implies that social evaluation is not only a function of identity but also a function of the properties of images, and so is changeable over time. Our social perception of individuals is flexible and dynamic. As both face- and voice-perception models suggest a somewhat independent processing of identity and emotion information in separate pathways, investigating social person evaluation can provide us with essential insight into the possible interaction between those pathways. Combining faces and voices together, therefore, can better inform our knowledge of both audio-visual integration and general models of face and voice processing, alongside bringing us closer to understanding integrated person perception.

References

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

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

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*, 711–725.

Berry, D. S. (1990). Vocal Attractiveness and vocal babyishness - effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior, 14*, 141–153.

Berry, D. S. (1991). Accuracy in social perception: contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*, 298–307.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.22.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour, 82*, 55-59.

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

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*(8), 1467–1485.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40*(1), 202–223.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*, 535-543. http://dx.doi.org/10.1016/j.tics.2007.10.001

Chen, D., Halberstam, Y., & Alan, C. L. (2016). Perceived Masculinity Predicts US Supreme Court Outcomes. *PloS One, 11*, e0164324.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audiovisual integration of emotion expression. *Brain Research, 1242*, 126–135.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42-45.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*, 289-311. http://dx.doi.org/10.1080/026999300378824

Dimos, K., Dick, L., & Dellwo, V. (2015). Perception of levels of emotion in speech prosody. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra-and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88*, 143-156.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1–11.

Fecher, N. (2015). *Praat pitch alteration script*. Department of Language and Linguistics, University of York. Script for Praat.

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

Green, K., Kuhl, P., & Meltzoff, A. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics*, *50*, 524-536.

Hess, U., Kappas, A., & Scherer, K. (1988). Multichannel communication of emotion: Synthetic signal production. In Scherer, K. (Ed.), *Facets of Emotion: Recent Research* (pp. 161-182). Hillsdale, NJ: Erlbaum.

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature, 21*, 406-427.

Hudson, T., De Jong, G., McDougall, K., Harrison, P., and Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. *In Proceedings of the 16th International Congress of Phonetic Science*, Saarbrücken: Germany, 1809-1812.

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

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

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*, 2698–2704.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica, 46*, 117-125.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*, R143–R145. doi:10.1016/j.cub.2010.12.033

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Lindh, J. (2006). Preliminary F0 statistics and forensic phonetics. *Proceedings of the 15th annual International Association of Forensic Phonetics and Acoustics conference*, Department of Linguistics, Göteborg University: Sweden.

Little, A. C., Burt, D. M., & Perrett, D. I. (2006). What is good is beautiful: Face preference reflects desired personality. *Personality and Individual Differences, 41*, 1107–1118. doi: 10.1016/j.paid.2006.04.015

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review, 3*, 215-221. doi:10.3758/BF03212421

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One, 9*, e90779.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Mehrabian, A., & Ferris, S.R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*, 248-252.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*, 855-869.

Ohala, J. J. (1982). The voice of dominance. *The Journal of the Acoustical Society of America, 72*, S66–S66.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*, 1-16.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: appearance-based trait inferences and voting. *Journal of Nonverbal Behaviour, 34*, 83–110. doi: 10.1007/s10919-009-0082-1

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

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The Measurement of Meaning.* Urbana: University of Illinois Press.

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,* 283-296.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28,* 340-344.

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

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

Schweinberger, S. R., Herholz, A., & Stief, V. (1997). Auditory long-term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology, 50A,* 498–517.

Schweinberger, S. R., Kloth, N., & Robertson, D. M. (2011). Hearing facial identities: Brain correlates of face–voice integration in person identification. *Cortex, 47,* 1026-1037. http://dx.doi.org/10.1080/17470210601063589

Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology, 60,* 1446-1456. http://dx.doi.org/10.1080/17470210601063589

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Summerfield, A. Q. (1979). Use of visual information in phonetic perception. *Phonetica, 36,* 314-331.

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

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different images of the same person. *Psychological Science, 25,* 1404-1417. doi: 10.1177/0956797614532474

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

Tomlin, R. J., Stevenage, S. V., & Hammond, S. (2016). Putting the pieces together: Revealing face–voice integration through the facial overshadowing effect. *Visual Cognition*, 1-15.

Traunmüller, H., and Eriksson, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*. Unpublished manuscript. [Available at http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf].

Tsankova, E., Krumhuber, E., Aubrey, A. J., Kappas, A., Möllering, G., Marshall, D., & Rosin, P. L. (2015). The Multi-modal nature of trustworthiness perception. In *AVSP* (pp. 147-152).

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception*, *45*, 946–463.

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance. *Human Communication Research*, *26*, 148-171.

Vukovic, J., Jones, B. C., Feinberg, D. R., DeBruine, L. M., Smith, F. G., Welling, L. L., & Little, A. C. (2011). Variation in perceptions of physical dominance and trustworthiness predicts individual differences in the effect of relationship context on women's preferences for masculine pitch in men's voices. *British Journal of Psychology, 102*, 37-48.

Wells, T. J., Dunn, A. K., Sergeant, M. J. T., & Davies, M. N. O. (2009). Multiple signals in human mate selection: A review and framework for integrating facial and vocal signals. *Journal of Evolutionary Psychology*, *7*, 111-139.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395–412. http://dx.doi.org/10.1037/0022-3514.37.3.395

Wilson, J., & Rule, N. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of trustworthiness. *Social Psychological and Personality Science, 7*, 331–338. doi: 10.1177/1948550615624142

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, *102*, 959-974.

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in cognitive sciences*, *17*, 263-271.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*, 1497-1517.

Zuckerman, M. & Driver, R. E. (1989). What sounds beautiful is good - the vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*, 67–82.