J Nonverbal Behav (2015) 39:355–370
DOI 10.1007/s10919-015-0214-8

ORIGINAL PAPER

Dominant Voices and Attractive Faces: The Contribution of Visual and Auditory Information to Integrated Person Impressions

Constantin Rezlescu[1,2] · Tegan Penton[3] · Vincent Walsh[4] · Hikaru Tsujimura[3] · Sophie K. Scott[4] · Michael J. Banissy[3,4]

Published online: 5 June 2015

Abstract Faces and voices, in isolation, prompt consistent social evaluations. However, most human interactions involve both seeing and talking with another person. Our main goal was to investigate how facial and vocal information are combined to reach an integrated person impression. In Study 1, we asked participants to rate faces and voices separately for perceived trustworthiness, attractiveness, and dominance. Most previous studies relied on stimuli in which extra-vocal information (e.g., verbal content, prosody) may have confounded voice-based effects; to prevent these unwanted influences, we used brief, neutral vowel sounds. Voices, like faces, led to the formation of highly reliable impressions. Voice trustworthiness correlated with voice attractiveness, mirroring the relation between face trustworthiness and attractiveness, but did not correlate with voice dominance. Inconsistent with the possibility that face and voice evaluations are indicative of real character traits, we found no positive correlations between judgments of trustworthiness or dominance based on faces and the same judgments based on voices (there was also no correlation between face attractiveness and voice attractiveness). In Study 2, we asked participants to evaluate male targets after seeing their faces and hearing their voices. Faces and voices contributed equally to judgments of trustworthiness and combined to produce a significant interaction effect. For attractiveness, faces were given more weight than voices, possibly due to the predominantly visual character of the attractiveness concept (there was no interaction effect). For dominance, the reverse pattern was true, with voices having a larger effect than faces on final judgments. In this case the auditory cues may be perceived to be more reliable because of the strong links between voice pitch, masculinity, and dominance.

✉ Constantin Rezlescu
rezlescu@fas.harvard.edu; rezlescu@gmail.com

[1] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA

[2] Department of Cognitive, Perceptual and Brain Sciences, University College London, London, UK

[3] Department of Psychology, Goldsmiths University, London, UK

[4] Institute of Cognitive Neuroscience, University College London, London, UK

Keywords Person perception · Trait judgments · Voice evaluations · Trustworthiness · Dominance

Introduction

First impressions may play an important role in social and economic interactions (Hamermesh 2011; Rezlescu et al. 2012; Todorov et al. 2005; Zebrowitz and McDonald 1991). Two of the critical cues used to derive first impressions are faces and voices, which are rich sources of socially relevant information. Looking at a face or hearing a voice we can reliably infer an individual's sex, age, identity, and emotional state (e.g., Banissy et al. 2010; Meyer et al. 2007; Sauter et al. 2010; Scott 2008). But faces and voices also prompt spontaneous evaluations related to attractiveness, and to character traits such as trustworthiness and dominance (Willis and Todorov 2006; Vukovic et al. 2011). These evaluations are highly consistent across observers (Oosterhof and Todorov 2008). Although a relatively large body of research has examined face- and voice-based impressions independently, there is less research investigating the interaction between the two. The main goal of the current study was to examine how face and voice information combine to form an integrated person impression.

We focused on evaluations of trustworthiness, attractiveness, and dominance for two reasons. First, an influential model of facial trait perception (Oosterhof and Todorov 2008) suggests that judgments of trustworthiness and dominance may account for the majority of variance in a large number of face-based judgments. The model is consistent with other two-dimensional models of social perception proposed previously, according to which most person impressions can be reduced to warmth and competence (Fiske et al. 2007) or affiliation and dominance (Wiggins et al. 1989). Second, attractiveness has been proposed to contribute to the overall accuracy of trait judgments. Previous studies suggested that participants may perform above chance when making face-based trait judgments (e.g., Jones et al. 2012; but see Olivola and Todorov 2010), and prior work suggests that attractiveness cues influence these judgments (e.g., Albright et al. 1988; Little and Perrett 2007).

Our study also aimed to improve the auditory stimuli used to study voice-based impressions. Most previous studies used sentences and words, raising the possibility that the observed effects were linked to extra-vocal information. Pear (1931) conducted one of the first studies on voice-based impressions through a British radio station. Four-thousand listeners estimated several characteristics of nine readers. The best predictions were about age and sex; a result which was later confirmed by Allport and Cantril (1934). These results, and many of the results from subsequent studies investigating voiced-based impressions, are confounded by cues such as semantics and intonation within speech (Kramer 1963). An early attempt to address such confounds comes from Scherer (1972) who used scrambled voice clips from longer conversations to demonstrate that peer, rather than self, ratings better correlate with viewer ratings of American and German target personality traits. While Scherer's splicing technique eliminated the semantics confound, it failed to control for emotional prosody, pitch, and volume. Other studies have used single spoken words (e.g., Aronovitch 1976; Berry 1991; McAleer et al. 2014) as a way to control for speech confounds while investigating voice impressions. However, even single spoken words provide information on top of basic voice information, such as accents, that can bias the trait evaluations. Our goal was to isolate voice information similarly to how face information is isolated when we look at photographs, and understand how the two types of information interact to provide trait cues.

In the literature, there have been several attempts to investigate the relative effects of voice-based and face-based information when the two are presented together. Most real-world interactions involve simultaneous access to facial and vocal information; therefore, it is important to understand the relative influences of each, and the interaction between the two. Zuckerman et al. (Zuckerman and Driver 1989; Zuckerman et al. 1990, 1991, 1995) have extensively investigated the relative effects of voice and face attractiveness on trait judgments. Specifically, Zuckerman and Driver (1989) observed that vocal attractiveness and facial attractiveness have similar influence on personality judgments (i.e., both led to more positive judgments). Furthermore, the effect of both modalities was reduced when engaging with both voice and face stimuli at once, highlighting the interaction between the two; a finding that has been replicated within the original study and by others (e.g., Berry 1990, 1992). Further work by Zuckerman and Colleagues (e.g., Zuckerman et al. 1990, 1991, 1995) has also demonstrated that manipulation of either voice or face attractiveness affects subsequent trait and attractiveness judgments (in the unaltered modality). This work provides robust evidence of the importance of both face and voice attractiveness during trait judgment. While the above research demonstrates robust findings related to speech, it does not allow for inferences to be made about the similarity of face and pure voice based impressions. Namely, due to accent, inflection, and emotional prosodies present in Zuckerman and Driver's (1989) voice stimuli, it is unclear to what extent voice alone has contributed. We used one-second clips of vowel annunciation [similar to those used by Collins (2000), Latinus and Belin (2012)] to control for the aforementioned confounds as well as to ensure that auditory and visual stimuli are comparable.

The current study aims to extend the findings of Zuckerman and Driver (1989), Zuckerman et al. (1990, 1991, 1995), Zuckerman and Sinicropi (2011) by investigating the importance of voice- and face- based cues when judging neutrally presented targets on three dimensions: attractiveness, dominance, and trustworthiness. In Study 1, we validate a novel set of neutral voice and face stimuli and examine the intramodal and intermodal correlations. In Study 2, we examine how face and voice evaluations combine to form an integrated person impression.

Study 1: Independent Evaluations of Voices and Faces

Method

Participants

We recruited online participants through Amazon mechanical turk. Results from web-based samples of participants have been shown to have comparable means, standard deviations, and internal reliability as those from lab-based samples (Berinsky et al. 2012; Buhrmester et al. 2011; Germine et al. 2012; Horton et al. 2011). Participants were US-based and had a task completion rate of at least 95 %, an indicator of good previous performance.

Two hundred ninety-eight mechanical turks (MTurks) rated the voice and face stimuli in exchange for $.60. Fifteen MTurks were excluded because of abnormally low variance in ratings in at least one block (cut-off threshold .7, corresponding to the bottom 5 % in a normal distribution with mean 3.3 and standard deviation 1.6 as in our sample; ratings were from 1 to 9). Therefore, we analyzed the ratings from 283 participants (age range 18–67, mean age 31.1, *SD* = 11.9; 141 female).

Voice and face stimuli. We obtained audiovisual material from 41 volunteers (age range 19–49, 19 female) recorded in a sound proof chamber. Volunteers were asked to face the camera and pronounce five English vowels lasting approximately one second each, in various expressions (neutral, happy, angry, surprised, disgusted, fearful and sad, in this order). Each vowel-expression was repeated once.

For the current study, we selected as voice stimuli the first neutral expressions of vowels A, E, and O (we discarded I and U because of the potential verbal loadings). Stimuli were normalized to a standard intensity level using Audacity (http://audacity.sourceforge.net). The face stimuli were full grayscale head shots in neutral expression. The volunteers were grouped into 'white British' (WB) (*n* = 26; nine female) and 'other' (OT) (*n* = 15; ten female) based on their reported ethnicity.

Table 1 Interrater agreements and reliabilities for independent social evaluations of voices and faces (italicised)

Trait judgment	Stimuli	Sample size (*n*)	White british targets (26)		Other targets (15)	
			Interrater agreement (*r*)	Reliability (α)	Interrater agreement (*r*)	Reliability (α)
Trustworthiness	Vowel A	30	**.29**	**.93**	.09	.76
	Vowel E	31	**.23**	**.90**	.08	.71
	Vowel O	40	**.25**	**.93**	.08	.79
	Faces	*101*	***.28***	***.97***	*.21*	*.97*
Attractiveness	Vowel A	30	**.24**	**.91**	.10	.77
	Vowel E	30	**.30**	**.93**	.10	.79
	Vowel O	29	**.33**	**.93**	.15	.83
	Faces	*89*	***.33***	***.98***	*.42*	*.98*
Dominance	Vowel A	29	**.22**	**.88**	.36	.93
	Vowel E	31	**.22**	**.90**	.26	.92
	Vowel O	33	**.22**	**.90**	.19	.88
	Faces	*93*	***.37***	***.98***	*.35*	*.98*

Stimuli were recorded from 26 white British individuals and 15 other individuals. Note that Cronbach's α tends to increase with sample size (number of raters); therefore the slightly higher reliabilities for face stimuli may simply be due to larger sample size. White British targets (in bold) were selected for subsequent analyses. Other targets were discarded because of lower consistency of judgments

Procedure

In a between-subjects design, MTurks were randomly assigned to one of nine [three trait evaluations (trustworthiness, attractiveness, dominance) $\times$ three vowels (A, E, O)] conditions. They were instructed to provide judgments on the assigned trait (e.g., attractiveness) for the 41 vocalizations of the assigned vowel as well as judgments on the same assigned trait for the 41 neutral face stimuli. Face stimuli did not differ according to vowels (i.e., the same faces were presented regardless of the vowel condition), thus the number of ratings per trait, for face stimuli, was approximately triple the number of ratings per vowel, per trait, for voice stimuli. The ratings ranged from 1 (*not at all trustworthy/attractive/dominant*) to 9 (*extremely trustworthy/attractive/dominant*). The trials were blocked as follows: WB voices, WB faces, OT voices, OT faces. The order within each block was randomized. The block order was fixed to allow discarding the stimuli groups not passing the validation criteria (we anticipated that the OT judgments would have lower reliability).

Results

Random allocation of MTurks to condition groups aimed to ensure the groups were comparable in terms of participant characteristics. Chi square values for 3-by-3 contingency tables (trait-by-vowel) confirmed the groups did not differ for total number of participants ($p = .900$), number of female participants ($p = .622$), or number of Caucasian participants ($p = .460$). A two-way ANOVA confirmed there was no difference in age between groups according to trait ($p = .921$) or vowel ($p = .695$), just a marginally significant interaction effect ($p = .046$, partial $\eta^2 = .035$), mainly driven by the slightly younger age ($M = 26.2$) of MTurks making trustworthiness judgments on vowel A.

Validation of voice and face stimuli. We first analyzed whether target stimuli were rated differently based on their ethnicity (WB or OT). Average ratings of trustworthiness and attractiveness were lower for WB voices than for OT voices (trustworthiness, 5.14 vs. 5.27, $t_{100} = 2.36$, $p = .020$; attractiveness, 4.62 vs. 4.85, $t_{88} = 2.90$, $p = .005$). WB faces were also perceived as more dominant than OT faces (4.92 vs. 4.73, $t_{92} = 2.68$, $p = .009$). Other ratings (dominance for voices, trustworthiness and attractiveness for faces) did not differ between WB and OT targets (all ps $> .05$).

As can be seen in Table 1, interrater agreements (computed as the averages of all pairwise correlations between raters' judgments) and reliabilities (Cronbach's α) of evaluations of both WB and OT faces were high, in line with previous reports (e.g., Oosterhof and Todorov 2008). Evaluations of WB voices were also highly reliable (above .88) and had interrater correlations comparable to faces (between .22 and .33). In contrast, evaluations of OT voices, especially the dominance judgments, were characterized by lower reliability and poor interrater agreement. For this reason, the OT stimuli (both faces and voices) were excluded from subsequent analyses.[1]

We next examined whether trait judgments across different vowels correlate, in other words, do individual voices receive similar ratings regardless of the vowel they are asked to produce? Correlations between judgments based on different vowels ranged between .62 and .76 (excluding the correlation between dominance judgments for the O and E vowels which was .47) and were significant across all categories (Table 2). This suggests trait

[1] Following one reviewer's suggestion, we also analyzed the OT data. Most of the results reported for WB stimuli hold for OT stimuli. We choose not to report/interpret them because the low reliability of judgments involving these stimuli precludes one from drawing firm conclusions.

Table 2 Correlations between social evaluations of same voices under different vocalizations (vowels A, E, and O)

Vowels	Trustworthiness	Attractiveness	Dominance
A–E	**.625 (.001)**	**.736 (.000)**	**.757 (.000)**
A–O	**.623 (.001)**	**.688 (.000)**	**.669 (.000)**
O–E	**.627 (.001)**	**.763 (.000)**	**.471 (.015)**
Mean correlation	*.625*	*.729*	*.632*

Average values are given in italics

Correlations were moderate to high and significant (in bold). *p* values are given in parentheses (values lower than .001 are presented as .000)

perceptions are linked to intrinsic voice properties. In the following sections, for simplification, we collapsed ratings of different vowels into one general rating per voice (and per trait) and report results based on these average voice ratings.

Intramodal Correlations

For male faces, trustworthiness judgments correlated positively with attractiveness judgments ($r_{15} = .51$, $p = .037$) and negatively with dominance judgments ($r_{15} = -.60$, $p = .012$). The correlation between dominance and attractiveness judgments was not significant ($r_{15} = -.21$, $p = .417$). For female faces, there were no significant correlations between trustworthiness, dominance and attractiveness judgments (see Table 3 for details).

For voices, trustworthiness and attractiveness judgments were highly correlated for both male and female targets ($r_{15} = .87$, $p < .001$; and $r_7 = .83$, $p = .005$, respectively). None of the other correlations reached significance (see Table 3).

Notably, the results did not suffer major changes when analyzed separately based on MTurks' gender.

To check whether the correlation coefficients for the male stimuli were significantly different from the correlation coefficients for the female stimuli, we used the Fisher r-to-z transformation and compared the obtained z score differences for statistical significance. None of the differences reached significance; even when the difference in the correlation coefficients was substantial, e.g., .014 versus .508 for female versus male faces Tw–At correlations, the *p* value did not approach significance ($p = .263$ for the above example). In this context, we acknowledge the limited power of our design (with nine female and 17 male stimuli) to detect statistically significant differences in the correlation coefficients according to the stimuli gender. With the same number of stimuli, the minimum difference between the correlation coefficients to be statistically significant is .74; for the maximum difference we obtained (.494), the minimum number of stimuli in each gender category for this difference to be significant is 29.

Intermodal Correlations

An initial comparison of face-based and voice-based judgments of targets' trustworthiness and attractiveness revealed significant correlations across stimuli types (all results are summarized in Table 4). However, when judgments were analyzed separately by target gender, these correlations were no longer significant. This suggests that significant correlations may have been due to the previously documented bias to judge feminine targets

Table 3 Correlations between social evaluations within stimuli type (voices and faces)

Social evaluations	Voices			Faces		
	All targets (n = 26)	Female targets (n = 9)	Male targets (n = 17)	All targets (n = 26)	Female targets (n = 9)	Male targets (n = 17)
Tw–At	**.890** (**.000**)	**.832** (**.005**)	**.865** (**.000**)	**.512** (**.008**)	.014 (.971)	**.508** (**.037**)
Tw–Do	.020 (.923)	.149 (.702)	.344 (.176)	−**.630** (**.001**)	−.410 (.273)	−**.595** (**.012**)
Do–At	−.334 (.095)	−.270 (.482)	.032 (.904)	−.282 (.163)	.349 (.358)	−.211 (.417)

There were 26 voice stimuli and 26 face stimuli, recorded from the same nine female and 17 male targets. Significant correlations are in bold and *p* values are given in parentheses (not corrected for multiple correlations). None of the correlation coefficients were significantly different for female and male targets (the lowest *p* value was .263 for the face Tw–At correlation coefficients; we used the Fisher r-to-z transformation to assess the significance of the difference between two independent correlations). Pattern of correlations was similar when results were analysed separately per participants' gender

Table 4 Correlations of social evaluations across stimuli type (voices and faces)

Social evaluations	White British targets		
	All targets (n = 26)	Female targets (n = 9)	Male targets (n = 17)
Trustworthiness	**.471** (**.015**)	.096 (.806)	.365 (.149)
Attractiveness	**.502** (**.009**)	.447 (.228)	.086 (.744)
Dominance	−.126 (.540)	−.026 (.947)	−**.515** (**.034**)

Significant correlations are in bold and *p* values are given in parentheses (not corrected for multiple correlations). Note that despite not reaching significance (perhaps due to reduced statistical power), attractiveness judgments for female voices and faces were moderately correlated. None of the correlation coefficients were significantly different for female and male targets. Pattern of correlations was similar when results were analysed separately per participants' gender

more favourably than masculine targets (Perrett et al. 1998). We note that, despite not reaching significance (perhaps due to reduced statistical power), attractiveness judgments of female voices and faces showed a modest correlation ($r_7 = .45$, $p = .228$).

Analysis of dominance judgments revealed an unexpected negative correlation between face and voice judgments for male targets in that faces perceived as being more dominant were coupled with voices perceived as being less dominant.

A separate analysis per participants' gender revealed the same pattern of correlations between evaluations of voices and faces.

The correlation coefficients obtained for the male stimuli were not significantly different from the correlation coefficients obtained for the female stimuli (*p* values .555, .418, and .267, for trustworthiness, attractiveness, and dominance judgments, respectively). Again, we note the limited power of our design to detect statistically significant differences between correlation coefficients corresponding to the two target gender (see results under Intramodal correlations).

Discussion

Intramodal Correlations

Research into trait evaluations from faces typically demonstrates a positive correlation between trustworthiness and attractiveness judgments, while a negative correlation is often observed between trustworthiness and dominance judgments (Oosterhof and Todorov 2008; Vecchiato et al. 2014). We replicated these effects for male faces, but not for female faces. These differences in perception of female and male faces underline the importance of dividing all analyses of trait perception by target gender.

The high correlations between trustworthiness and attractiveness judgments (for male faces and voices, and female voices) may imply that participants, in general, find it difficult to judge these two dimensions independently, possibly due to the use of highly standardized stimuli (i.e., neutral, grey-scaled images and short, volume- and content-controlled voice clips). This may result in the coupling of dimensions, leading individuals to judge faces similarly on all positive (and negative) attributes (i.e., the halo effect; Dion et al. 1972). Furthermore, the null correlation between trustworthiness and attractiveness observed in face judgments for female targets suggests that when relying on facial cues, this effect may be specific to target sex.

Seminal evidence from Buss et al. (1990) suggests that attractiveness and trustworthiness traits are seen as more desirable in women, while traits related to wealth (i.e., good earning capacity) are seen as more desirable in men. Therefore, it may be that viewers have developed mechanisms to enable them to better identify cues related to both attractiveness and trustworthiness in females, both for mate selection (in male viewers) and for competition (in female viewers). Thus, when viewing male faces, people may be less interested in identifying visual cues discriminating between these constructs and provide instead a more general valence judgment incorporating both attractiveness and trustworthiness (and possibly other traits).

Note that the above interpretation of the differences in social evaluations of female versus male stimuli is based on the statistical significance of individual correlation coefficients (compared to zero correlation) and not on direct contrasts between coefficient correlations for female versus male stimuli. None of these contrasts were significant, but our design had limited statistical power for this purpose.

For female faces, trustworthiness judgments seem to be orthogonal to attractiveness judgments (the correlation between the two is virtually zero). However, for female voices, trustworthiness and attractiveness judgments are highly correlated. A potential explanation for this striking difference between face-based and voice-based evaluations may be that people find the attractiveness and/or trustworthiness concepts more ambiguous when applied to auditory stimuli, and therefore tend to provide general valence evaluations instead of specific trait judgments.

Intermodal Correlations

There are two reasons to be interested in looking at intermodal correlations of trait judgments. First, because people agree to a large extent with respect to which faces (or voices) are high on a given trait (e.g., Albright et al. 1988; Bar et al. 2006; Ekman et al. 1980), it is reasonable to assume that there are certain visual (or auditory) cues which are consistently seen as signalling that particular trait. It is of interest to investigate whether the same persons display both the telling visual and auditory cues. The second reason is an extension of the first. It has been suggested that certain character traits (including trustworthiness) can be accurately inferred from faces and voices (e.g., Ambady et al. 1995; Berry, 1990, 1992; Carré and McCormick 2008; Stirrat and Perrett 2010). If this were the case, one may expect some diagnostic information to be gained through both faces and voices.

In general, we did not find evidence for commonalities between face- and voice-based judgments. The only exception was the dominance judgments of male faces and voices, between which we found a surprising negative correlation, suggesting that dominant faces had less dominant voices. It has been suggested that male facial width ratio (Verdonck et al. 1999) and voice pitch (Dabbs and Mallinger 1999) are both influenced by testosterone level, which in turn has been linked with dominant and aggressive behavior (Ehrenkranz et al. 1974). Therefore, a positive correlation between ratings of voice and face dominance was expected, reasoning that men with wider faces tend to have lower-pitched voices and both wide faces and low-pitched voices would be perceived as dominant. The negative correlation may be tentatively explained in terms of mate selection and competition: men with less dominant facial features may need to compensate with more dominant voices, and vice versa. Research by Fraccaro et al. (2013) shows that deliberate lowering of pitch in men's voices leads viewers to perceive them as more dominant but not more attractive. This would suggest that this strategy is largely employed for intra-sex competition with less emphasis on mate selection. At the same time, males with dominant faces may wish to be perceived as less dominant by females. While dominance is seen as evolutionarily desirable in the male species to cope with competition from other males, some studies suggest that if a male appears too dominant, he becomes less attractive as a long-term mate to a female (Johnston 2006; Valentine et al. 2014). Research by O'Connor et al. (2011) showed that men with more masculinized voices (lower pitch) are judged as more likely to be unfaithful than those with less masculinized voices. Therefore, males may use their voices to achieve the fine balance between being perceived as dominant enough to fend off competition and still appearing attractive to the opposite sex.

Study 2: Integrated Person Perception

As we have seen, people tend to associate positive or negative traits with voices, similar to processes taking place with faces (e.g., Berry 1990, 1992; Zuckerman and Driver 1989). Face evaluations were shown to play a significant role in social interactions (e.g., Hamermesh 2011) even when reputational information is available (Rezlescu et al. 2012), and voice evaluations have a similar impact (e.g., Berry 1991; Pear 1931). However, social interactions typically involve simultaneous access to facial and vocal information, so a person impression is likely to be based on both the face and the voice [as seen in work by Zuckerman et al. (1990, 1991, 1995, Zuckerman and Sinicropi (2011)]. Our next aim was to examine how face evaluations and voice evaluations combine to form an integrated person impression.

Method

Participants

There were three groups of online participants (MTurks) corresponding to the three trait judgments. The 'trustworthiness' group had 41 participants (age range 18–59, mean age

30.9, $SD = 9.8$; 17 female), the 'attractiveness' group had 41 participants (age range 19–66, mean age 34.0, $SD = 13.4$; 18 female) and the 'dominance' group had 84 participants (age range 18–51, mean age 28.6, $SD = 8.2$; 34 female). The considerably larger number of participants in the 'dominance' group was determined by a script error that doubled the requested number of participants compared with the other groups (the results did not differ when keeping only the first 41 participants). The task lasted 2 min and each participant received $.25.

Voice and Face Stimuli

This experiment used a sample of the male WB stimuli ($n = 17$) from Study 1. Female stimuli were excluded to control sex-related biases (e.g., male faces and voices may be evaluated differently when presented together with female faces and voices). The voice stimuli were of vowel A. After excluding targets above 40 years old ($n = 3$), with mild strabismus ($n = 1$) or facial hair ($n = 1$), we were left with 12 targets whose faces and voices could be mixed.

Procedure

First, faces were ordered based on their trustworthiness ratings from Study 1 and binned accordingly into two equal groups, one with 'untrustworthy' (UTw) and the other with 'trustworthy' (Tw) faces. We then did the same for voices, producing one group of UTw and one group of Tw voices. Note that the face and voice of a particular target could fall in different categories (e.g., one target could have a 'trustworthy' face and an 'untrustworthy' voice). Each group had six face (voice) stimuli. The average ratings for the four categories were: 3.9 for UTw faces, 5.1 for Tw faces, 4.4 for UTw voices and 5.3 for Tw voices. For each participant, we randomly paired half of the untrustworthy faces with untrustworthy voices and the other half of the untrustworthy faces with trustworthy voices. Similarly, half of the trustworthy faces were paired with untrustworthy voices and the other half with trustworthy voices. In the end, we had four experimental conditions, with three pairs per condition. After a catch trial ensuring that that they have the sound on, participants were presented each pair (face plus voice) in a randomized order and asked to rate the respective person for trustworthiness.

The same procedure was used to create the pairs of stimuli and collect ratings for the attractiveness and dominance judgments. The average ratings for the respective categories were: 2.9 for 'unattractive' (UAtt) faces, 4.0 for 'attractive' (Att) faces, 3.9 for Uatt voices, 5.0 for Att voices; 4.3 for 'undominant' (UDom) faces, 5.8 for 'dominant' (Dom) faces, 4.4 for UDom voices, and 5.6 for Dom voices. Allocation of participants to the trustworthiness, attractiveness or dominance condition was counterbalanced.

Results

The pattern of results for the three trait judgments can be seen in Fig. 1 and is summarized in Table 5.

The significance thresholds for the p values reported below were adjusted to reflect the number of factors in each ANOVA. Thus, for two-way ANOVAs (two main effects and one interaction), only p values below .017 were considered significant. For three-way

ANOVAs (seven factors: three main effects, three two-way interaction, and one three-way interaction), the significance threshold was .007.

We calculated an average rating per condition per participant and submitted these numbers to three 2 × 2 (Face × Voice) within-subjects ANOVA, one per trait (Fig. 1). For trustworthiness judgments, both main effects of face and voice were significant, face $F_{1,40} = 25.06$, $p < .001$, partial $\eta^2 = .39$; voice $F_{1,40} = 30.85$, $p < .001$, partial $\eta^2 = .44$. In addition, there was a significant interaction effect between face and voice trustworthiness, $F_{1,40} = 9.67$, $p = .003$, partial $\eta^2 = .20$, with a larger impact of Tw faces on Tw voices than on UTw voices.

The main effects of face and voice were also significant for attractiveness judgments; face $F_{1,40} = 56.08$, $p < .001$, partial $\eta^2 = .58$; voice $F_{1,40} = 6.37$, $p = .016$, partial $\eta^2 = .14$. There was no interaction effect between faces and voices for attractiveness ($p = .086$). We note that the main effect size of faces is four times the main effect size of voices. This difference could not be attributed to a larger difference in the average ratings of Att and UAtt faces versus the difference in the average ratings of Att and UAtt voices, because they were identical (see "Procedure" section).

For dominance judgments, again both faces and voices had significant main effects; faces $F_{1,83} = 16.39$, $p < .001$, partial $\eta^2 = .17$; voice $F_{1,83} = 91.61$, $p < .001$, partial $\eta^2 = .53$. The interaction effect between faces and voices was not significant: $F_{1,83} = 3.57$, $p = .062$, partial $\eta^2 = .04$. We note that in terms of effect sizes, the situation is reversed compared to attractiveness judgments; for dominance, voices have an effect size three times larger than faces.

Transforming MTurks' ratings into z scores and reanalyzing the data had no impact on the reported results. Similarly, adding rater gender as a between-subjects factor to the above ANOVAs produced negligible changes. In all cases, main effects of gender were not significant (all $ps > .49$) and there were no interaction effects with the other factors (all $ps > .13$, with the exception of the marginally significant triple interaction effect between faces and voices for dominance judgments, $p = .032$, denoting that the interaction effect between faces and voices tended to be more pronounced for male raters).

To confirm the differential influence of faces and voices on judgments of attractiveness and dominance, we ran a three-way mixed ANOVA with face and voice as within-subjects

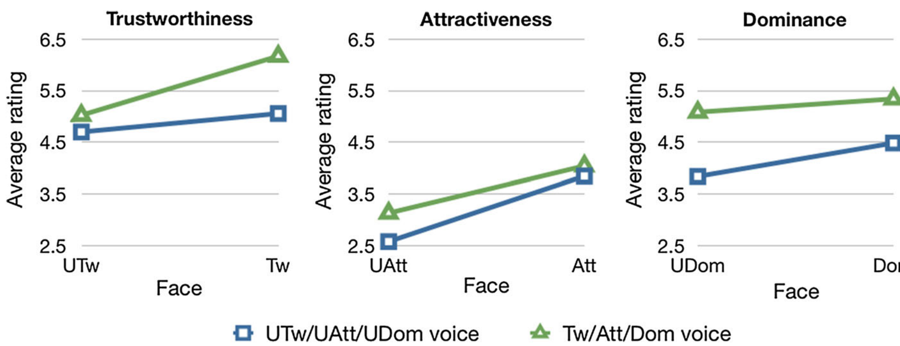

Fig. 1 Average ratings to composite stimuli (faces and voices). Voices had a higher impact on judgments of dominance than on judgments of attractiveness, as evidenced by the larger gap between the *green line* (voices with positive valence) and the *blue line* (voices with negative valence). Faces had a higher impact on attractiveness than dominance evaluations, as evidenced by the steeper slopes in the *blue* and *green lines* (Color figure online)

Table 5 Effect sizes (partial η^2) for faces and voices in forming person impressions as revealed by a two-way factorial ANOVA

Trait	Main effects			Interaction
	Face	Voice	Ratio (rounded)	Face × voice
Trustworthiness	.39	.44	1/1	.20
Attractiveness	.58	.14	4/1	–
Dominance	.17	.53	1/3	–

Faces and voices had an equal contribution (and significantly interacted) in trustworthiness judgments. Faces were more important for attractiveness judgments, while voices were more important for dominance judgments. Only significant effects are shown

factors, and trait (attractiveness/dominance) as between-subjects factor. All main effects were significant (face $F_{1,123} = 66.15$, $p < .001$, partial $\eta^2 = .35$; voice $F_{1,123} = 57.30$, $p < .001$, partial $\eta^2 = .32$; trait: $F_{1,123} = 46.23$, $p < .001$, partial $\eta^2 = .27$). Most importantly for our purposes, the two-way interaction Face × Trait was significant ($F_{1,123} = 11.29$, $p = .001$, partial $\eta^2 = .08$), reflecting a higher impact of face on evaluations of attractiveness than on evaluations of dominance. Similarly, the two-way interaction Voice × Trait was also significant ($F_{1,123} = 12.91$, $p < .001$, partial $\eta^2 = .10$), reflecting a higher impact of voice on evaluations of dominance than on evaluations of attractiveness. The interaction between faces and voices approached significance ($F_{1,123} = 5.20$, $p = .024$, partial $\eta^2 = .04$), and there was no triple interaction effect ($F_{1,123} = .01$).

Discussion

Our results show that integration of face and voice information in forming person impressions differs according to the judged trait. Thus, while for all measured traits both faces and voices have a significant impact on person impressions, the effect sizes are of different magnitudes. Faces seem to be more important than voices when it comes to attractiveness evidenced by a larger effect size for the former, while the reverse pattern was observed for dominance judgments. Trustworthiness impressions were equally influenced by faces and voices. Additionally, we found a significant interaction effect between faces and voices for trustworthiness judgments: the impact of voice trustworthiness was higher on trustworthy faces (increase in average rating 1.1) than on untrustworthy faces (increase on average rating .3). The different pattern of visual and auditory information integration for trustworthiness and attractiveness judgments suggests that, despite a high correlation between these judgments within the same modality, they are still distinct judgments. In other words, when asked to judge perceived trustworthiness, people rely on more than just attractiveness.

For attractiveness, our findings are partly consistent with a previous study by Surawski and Ossoff (2006) who found a higher impact of visual cues related to attractiveness on character trait judgments of political candidates. In that study, the effects of face attractiveness and voice attractiveness on perceived competence, trustworthiness and leadership were assessed. The results showed a visual superiority effect and a synergistic combination of face and voice attractiveness on evaluations of other character traits of political candidates. The higher impact of visual information was evident in our study, but we did not find a significant interaction between faces and voices in evaluations of person attractiveness. The difference may be due to different task demands; while we measured integrated attractiveness, Surawski and Ossoff (2006) did not. As we have seen in the present study, people integrate information differently according to the judged trait.

Previous studies on bimodal integration report a dominant role for visual information across multiple domains (e.g., Riggio et al. 1991; Busso et al. 2004). Interestingly though, this visual superiority effect varies according to signal reliability. For example, Collignon et al. (2008) showed that visual dominance in emotional processing disappeared when the reliability of the visual stimuli was diminished. In our study, the different contributions of faces and voices to attractiveness and dominance perception may be explained within this framework. On the one hand, if we consider that attractiveness is predominantly a visual concept, visual information should be considered most reliable and thus, the key determinant in overall judgments of attractiveness. On the other hand, dominance judgments have been shown to correlate with more masculine aspects (Oosterhof and Todorov 2008). Because the voice is highly sexually dimorphic (Fitch and Giedd 1999), it may be considered more reliable when judging a person's masculinity and therefore given higher weight in overall assessments of dominance. The voice pitch may even be a valid cue to dominance as it can signal testosterone level, and testosterone level is directly linked to actual dominance in men (Mazur and Booth 1998). Further, it may also be possible to identify more transient dominant states through vocal information. For example, Fraccaro et al. (2013) reported that manipulation of voice pitch influences judgments related to dominance. They argued that we may employ strategies to maximize social gain, in that attending to these variable cues when assessing dominance may reduce the likelihood of encountering conflict. Therefore, facial cues signaling dominance may be indicative of trait dominance (e.g., strong jawline indicative of high levels of testosterone) while voice cues signaling dominance may be indicative of state dominance (i.e., situation dependent). In line with this, Puts et al. (2006) showed that men who feel they are more dominant than their competitor will lower their voice pitch while the opposite pattern was observed in men who felt they were less dominant.

Conclusions

Our findings suggest that, while sharing some aspects with face impressions, voice evaluations also display specific properties. First, voices, just like faces, can lead to formation of consistent trait impressions of trustworthiness, attractiveness, and dominance. Our measures of reliability and interrater agreement were high and comparable to those of face impressions reported in the current study (e.g., Zuckerman and Hodgins 1993; Zuckerman et al. 1990). Second, we showed that voice-based judgments of trustworthiness and attractiveness were highly correlated, perhaps due to a valence halo effect used to compensate for the mismatch between the domain of the attractiveness judgment (predominantly visual) and the domain of the stimuli (auditory). Third, we showed that voice evaluations do not correlate with face evaluations (with the exception of perceived dominance for male stimuli). This suggests visual cues are not paired with auditory cues (diagnostic or not) when it comes to trait perception.

For face impressions, trustworthiness judgments of male targets correlated positively with attractiveness judgments and negatively with dominance judgments. In contrast, judgments of female targets did not correlate with each other. As we noted, a separate analysis per gender is necessary to avoid artificial correlations due to sex biases in trait perception. For example, if women are generally perceived as more trustworthy and more attractive than men, trustworthiness and attractiveness ratings will erroneously appear as correlated because of this divide between men and women.

Interpersonal perceptions in humans usually involve a combination of visual and auditory stimuli. In our second study, we showed that an integrated person impression assigns different weights to facial and vocal information, depending on the trait. Voices and faces are equally important to an integrated trustworthiness impression, which also shows an interaction effect. In contrast, attractiveness impressions seem to rely more on faces, while dominance impressions seem to rely more on voices. The differential roles of faces and voices in attractiveness and dominance evaluations may be explained by the predominantly visual character of attractiveness, on one hand, and by the strong links between voice pitch, masculinity and dominance, on the other hand.

In summary, our findings suggest that vocal and facial information differentially contribute to social evaluations dependent on the judgment. More complex judgments (e.g., trustworthiness) appear to require equal information from faces and voices, perhaps indicative of experienced difficulty of the viewer. For other judgments, such as attractiveness and dominance, the different weights assigned to visual and auditory information are likely related to the importance of each modality to the construct on which people are being evaluated, and may reflect strategic choices used to facilitate successful social evaluations.

Acknowledgments C.R. was supported by a Ph.D. fellowship from Axa Research Fund. M.J.B. was supported by the British Academy (PF100123) and ESRC (ES/K00882X/1).

References

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*(3), 387.

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

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*(3), 518–529.

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology, 99*, 207–220.

Banissy, M. J., Sauter, D. A., Ward, J., Warren, J. E., Walsh, V., & Scott, S. K. (2010). Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. *Journal of Neuroscience, 30*(41), 13552–13557.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis, 20*(3), 351–368.

Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior, 14*(3), 141–153.

Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*(2), 298–307.

Berry, D. S. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior, 16*(1), 41–54.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5.

Buss, D. M., Abbott, M., Angleitner, A., Asherian, A., Biaggio, A., Blanco-Villasenor, A., et al. (1990). International preferences in selecting mates A study of 37 cultures. *Journal of Cross-Cultural Psychology, 21*(1), 5–47.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., et al. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on multimodal interfaces* (pp. 205–211). ACM.

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences, 275*(1651), 2651–2656.

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

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60*(6), 773–780.

Dabbs, J. M., Jr., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*, 801–804.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285–290.

Ehrenkranz, J., Bliss, E., & Sheard, M. H. (1974). Plasma testosterone: Correlation with aggressive behavior and social dominance in man. *Psychosomatic Medicine, 36*(6), 469–475.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology, 38*(2), 270.

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

Fitch, T. W., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*, 1511–1522.

Fraccaro, P. J., O'Connor, J. J., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour, 85*(1), 127–136.

Germine, L., Nakayama, K., Duchaine, B., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857.

Hamermesh, D. S. (2011). *Beauty pays: Why attractive people are more successful*. Princeton: Princeton University Press.

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

Johnston, V. S. (2006). Mate choice decisions: The role of facial beauty. *Trends in Cognitive Sciences, 10*(1), 9–13.

Jones, A., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1353–1361.

Kramer, E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin, 60*(4), 408.

Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief vowel stimuli. *PLoS ONE, 7*(7), e41384.

Little, A. C., & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology, 98*(1), 111–126.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(3), 353–363.

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

Meyer, M., Baumann, S., Wildgruber, D., & Alter, K. (2007). How the brain laughs. Comparative evidence from behavioral, electrophysiological and neuroimaging studies in human and monkey. *Behavioural Brain Research, 182*(2), 245–260.

O'Connor, O., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology, 9*(1), 64–78.

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*(2), 83–110.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105*(32), 11087–11092.

Pear, T. H. (1931). *Voice and personality*. London: Chapman and Hall.

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

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

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS One, 7*(3), e34293.

Riggio, R. E., Widaman, K. F., Tucker, J. S., & Salinas, C. (1991). Beauty is more than skin deep: Components of attractiveness. *Basic and Applied Social Psychology, 12*(4), 423–439.

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 of the United States of America, 107*(6), 2408–2412.

Scherer, K. R. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality, 40*(2), 191–210.

Scott, S. K. (2008). Voice processing in monkey and human brains. *Trends in Cognitive Sciences, 12*(9), 323–325.

Stirrat, M. R., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349–354.

Surawski, M. K., & Ossoff, E. P. (2006). The effects of physical and vocal attractiveness on impression formation of politicians. *Current Psychology, 25*(1), 15–27.

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

Valentine, K. A., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a man by the width of his face: The role of facial ratios and dominance in mate choice at speed-dating events. *Psychological Science, 25*(3), 806–811.

Vecchiato, G., Toppi, J., Maglione, A. G., Olejarczyk, E., Astolfi, L., Mattia, D., et al. (2014). Neuroelectrical correlates of trustworthiness and dominance judgments related to the observation of political candidates. *Computational and Mathematical Methods in Medicine, 2014*, 434296. doi:10.1155/2014/434296.

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics, 21*(2), 137–143.

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*(1), 37–48.

Wiggins, J. S., Phillips, N., & Trapnell, P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology, 56*(2), 296–305.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598.

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*(6), 603–623.

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*(2), 67–82.

Zuckerman, M., & Hodgins, H. S. (1993). Developmental changes in the effects of the physical and vocal attractiveness stereotypes. *Journal of Research in Personality, 27*, 349–364.

Zuckerman, M., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype: Replication and elaboration. *Journal of Nonverbal Behavior, 14*(2), 97–112.

Zuckerman, M., Miyake, K., & Elkin, C. S. (1995). Effects of attractiveness and maturity of face and voice on interpersonal impressions. *Journal of Research in Personality, 29*, 253–272.

Zuckerman, M., Miyake, K., & Hodgins, H. S. (1991). Cross-channel effects of vocal and physical attractiveness and their implications for interpersonal perception. *Journal of Personality and Social Psychology, 60*(4), 545–554.

Zuckerman, M., & Sinicropi, V. (2011). When physical and vocal attractiveness differ: Effects on favorability of interpersonal impressions. *Journal of Nonverbal Behavior, 35*(2), 75–86.