Abstract. Research on human attraction frequently makes use of single-modality stimuli such as neutral-expression facial photographs as proxy indicators of an individual’s attractiveness. However, we know little about how judgments of these single-modality stimuli correspond to judgments of stimuli that incorporate multi-modal cues of face, body and speech. In the present study, ratings of attractiveness judged from videos of participants introducing themselves were independently predicted by judgments of the participant’s facial attractiveness (a neutral-expression facial photograph masked to conceal the hairstyle), body attractiveness (a photograph of the upper body), and speech attractiveness (the soundtrack to the video). We also found that ratings of the face, body and speech were positively related to each other. Our results support the assumption that the single-modality stimuli used in much attractiveness research are valid proxy indicators of overall attractiveness in ecologically valid contexts, and complement literature showing cross-modality concordance of trait attractiveness, but also recommend that research relying on assessments of individual attractiveness take account of both visual and vocal attractiveness where possible.

Keywords: facial attractiveness, body attractiveness, vocal attractiveness, attractiveness judgments

Studies of human attractiveness have found that: a) humans the world over are remarkably concordant in their judgments of the physical attractiveness of others (Langlois et al. 2000); b) individuals judged relatively attractive in one trait are often judged relatively attractive in another (e.g. Thornhill and Grammer 1999;
PETERS, RHODES and SIMMONS 2007); c) attractiveness judgments are context-dependent, becoming more attuned as they become more biologically relevant, such as at puberty (SAXTON, CARYL and ROBERTS 2006; SAXTON et al. in review) or at the peak fertility phase of a woman’s menstrual cycle (review in JONES et al. 2008); and d) physical attractiveness is associated with enhanced reproductive success (e.g. RHODES, SIMMONS and PETERS 2005; APICELLA, FEINBERG and MARLOWE 2007). These findings have been interpreted as evidence that judgments of physical attractiveness in humans are underpinned by particular psychological adaptations, and explain why physical attractiveness continues to be the subject of much research interest (review in ROBERTS and LITTLE 2008).

Although judgments of physical attractiveness can be made with reference to a wide range of physical attributes, most work on human attraction has employed single-modality stimuli such as photographs of faces or recordings of voices, with ratings of these traits often assumed to be a proxy for the overall attractiveness of an individual (review in RHODES 2006). Research has shown that attractiveness of both face and voice (ZUCKERMAN, MIYAKE and HODGINS 1991), face and body (MUESER et al. 1984; BROWN, CASH and NOLES 1986; RAINES, HECHTMAN and ROSENTHAL 1990a; HÖNEKOPP et al. 2007; PETERS, RHODES and SIMMONS 2007), and face and dynamic expressiveness (RIGGIO et al. 1991) can each contribute individually to perceptions of an individual’s attractiveness. Yet work on the contributions of various modalities (in particular, the relative importance of face and voice) to social impression formation (e.g. perceptions of dominance or trustworthiness) has emphasised that the contribution of each modality is dependent both upon the context in which judgments are made and the context in which stimuli are recorded (e.g. FRIEDMAN 1978; DEPAULO, ZUCKERMAN and ROSENTHAL 1980; EKMAN et al. 1980; NOLLER 1985; ZEBROWITZ-MCArTHUR and MONTEPARE 1989; RAINES, HECHTMAN and ROSENTHAL 1990a, b). To understand how the different modalities contribute to perceptions of attractiveness at zero acquaintance, then, research specific to judgments of attractiveness in zero acquaintance social contexts is necessary. Such research has shown that the face is more important than the rest of the body to judgments of static whole-body attractiveness (PETERS, RHODES and SIMMONS 2007), and that facial attractiveness and dynamic expressiveness are more important than attractiveness of body or dress (RIGGIO et al. 1991). Yet an examination of the various contributors to overall personal attractiveness in more naturalistic conditions, and in particular a consideration of the important modulatory effect of vocal attractiveness in overall attractiveness judgments (ZUCKERMAN, MIYAKE and HODGINS 1991), has so far been lacking.

Related to the question of the contributions of the different modalities to attractiveness judgments, there is a controversy over whether different traits each provide similar information on an individual’s attractiveness. Correlations between attractiveness ratings are predicted on the theoretical footing that physical traits may provide concordant information about an individual’s suitability as a partner (MÖLLER and POMIANKOWSKI 1993). In line with this, ratings of faces and odour
(RIKOWSKI and GRAMMER 1999) and male personality and odour (HAVLICEK, ROBERTS and FLEGR 2005) are positively correlated within an individual. Yet other findings have been mixed. Research has found a relationship in women but not in men between the ratings of faces and bodies (THORNHILL and GRAMMER 1999; PETERS, RHODES and SIMMONS 2007), and between faces and facial movement (MORRISON et al. 2007). Reports of the correspondence between attractiveness ratings of facial photographs and silent videos of the moving faces have reported variously that they correspond in stimuli of both sexes (BROWN, CASH and NOLES 1986; ROBERTS et al. in press), in female but not male stimuli (LANDER 2008; PENTON-VOAK and CHANG 2008), and not in female stimuli (RUBENSTEIN 2005). Some studies have linked facial or physical and vocal attractiveness (FEINBERG et al. 2005; SAXTON, CARYL and ROBERTS 2006) while others have found no such link (OGUCHI and KIKUCHI 1997; LANDER 2008). Similarly, correlations between attractiveness ratings of voices and video recordings of the faces have been reported as lacking (RAINES, HECHTMAN and ROSENTHAL 1990a) or as present only from opposite-sex judgments (LANDER 2008). Finally, no significant correlations have been found between attractiveness of face, body and dynamic movement (RIGGIO et al. 1991), and between ratings from photographs and ratings by peers (KNIFFIN and WILSON 2004). Such inconsistencies cast doubt on the validity of using single measures such as static facial photographs to serve as proxies for overall individual attractiveness.

Here, we explicitly test the contribution of, and relationship between, specific components in different modalities (i.e. face, body, and speech) to judgments of attractiveness. We first video recorded a set of target individuals as they introduced themselves. We then asked one group of raters to rate each of these introductions for attractiveness, and a second, independent group of raters to rate the various components (face, body, speech) of these videos for attractiveness. We investigated the relative importance of each of the isolated components to the overall rating, and whether there were correlations between the individual components.

**METHODS**

Fifty individuals (aged 18–39; mean ± SD 24 ± 4 years; 25 males) were recruited from the student population and from social contacts. Individuals provided informed consent to participate in a study on the contribution of different components to impression formation including initial attractiveness judgments. The study was approved by the University of Liverpool Committee on Research Ethics. Individuals were asked to introduce themselves as though meeting someone new, and to speak for at least 20 s. Sitters were recorded at a distance of approximately 1.5 m from the video camera (Sony DVD DCR200E) while seated in front of a white backdrop. The zoom was adjusted manually to frame the participant from top of head to mid-thigh. White balance, exposure and focus were manually set and held.
constant. Sitters were asked to introduce themselves as they might upon meeting someone for the first time in a pub or bar. They were prompted to talk on any subject they wished, including their studies, hobbies, work, free time, holidays or weekend activities. Sitters were free to request that the researcher leave the room for the duration of the recording, and to have a repeat take of the recording, for example if they failed to speak for 20 s. Sitters were photographed (Canon Powershot) in front of the white backdrop as they adopted a neutral expression looking straight at the camera. Seated photographs of the face alone, and then standing photographs of the entire face and body with arms held vertical and parallel with their sides, were taken. Individuals were also recorded (M-Audio Microtrack 24/96, Audio-Technica ATR55 Telemike Shotgun cardioid condenser microphone) as they read out a sentence: “The quick brown fox jumped over the lazy dog”. All recordings were made in the same location within a windowless recording room lit with standard fluorescent lighting. Facial photographs were cropped with an oval around the face outline, concealing the majority of the hair (Corel Paint Shop Pro Photo X2) and normalised in size with reference to interpupillary distance (c.f. BURRISS and LITTLE 2006), and body photographs were cropped from neck to fingertips (i.e. omitting the head) and normalised for height. Size normalisation was carried out with specialist image software (Psychomorph; TIDDEMAN, BURT and PERRETT 2001). Resultant image dimensions were 1276 x 1276 pixels (face) and 1181 x 1259 pixels (body). Video recordings were edited to a duration of 20 s, cropped to dimensions of 400 x 480 square pixels and encoded as 25 f.p.s. QuickTime movies using the MPEG-4 codec (Adobe After Effects 7.0). Soundtracks were normalised for amplitude to ensure volume similarity between recordings (Adobe Audition 2.0).

Raters (aged 18 – 32; mean ± SD 20 ± 3 years) were recruited from the student population of a different university from that at which sitters were recruited, and participated in one of three separate sessions in exchange for course credit. Participants filled out basic demographic information in an anonymous questionnaire. In two of the sessions, raters (N = 26; 7 males) were presented with the complete 20 s video of each self-introduction. Raters reported themselves to be white European, except for one who did not answer, and one rater reported herself to be homosexual while the rest reported themselves to be heterosexual. In a third session, raters (N = 13; 3 males) were presented with masked neutral face photographs, upper body images, and finally video soundtracks. Small numbers of raters are sufficient for judgments of attractiveness because attractiveness judgments tend to be homogeneous between raters (e.g. FEINBERG et al. 2005; JONES et al. 2005). Raters from the third session also rated a subset (N = 39) of the voice recordings of the neutral sentence. Female stimuli were presented before male stimuli in each case. Raters reported themselves to be white European, except for one who described herself as Asian, and one rater reported herself to be bisexual while the rest reported themselves to be heterosexual. Sessions took place in a classroom where stimuli were projected onto a screen and sounds played over a speaker system. Raters rated each
stimulus for attractiveness on a scale from 1–7 (anchors “unattractive” and “very attractive”) on a sheet of paper. They were told that the sitters were recruited from a normal student population, and with reference to this they should endeavour to use the whole rating scale if possible. They were instructed not to respond in any way which could allow others in the room to infer their judgments. The mean attractiveness rating of each component for each sitter was calculated.

Analysis was carried out using SPSS 15.0. Kendall’s coefficient of concordance ($W$) was used to examine concordance. One rater was confused at the beginning of the session and did not rate the first 25 masked faces, and so her ratings were excluded from the concordance calculations for all of the masked faces. If raters neglected to rate a stimulus ($N = 55$ of a possible 3225 ratings) values were replaced by the mean value given to that stimulus so that the rest of the rater’s ratings could be included in the analysis. Kendall’s tau-b ($\tau$) is used for correlational analysis because some sets of ratings were non-normally distributed (Shapiro-Wilk’s test of normal distribution, $p < .1$) once ratings had been split for separate analysis for male and female stimuli. Assumptions of regression were satisfied, including that of no multicollinearity, as measured with reference to correlation between pairs of variables, VIF, tolerance, loadings and determinant of the correlation matrix (Field 2005). Ratings of the neutral sentence were not included in the regression analysis because they correlated highly with ratings of the soundtrack ($r = .660$, $p < .001$). Substitution in the main analysis of the neutral sentence ratings for the soundtrack ratings gave rise to qualitatively identical results; such substitution was not carried out in the separate regressions for male and female targets because of the restrictions on sample size ($n = 39$ for whom sentence ratings were obtained). The enter method of regression was used because no assumption was made as to the relative importance of each component to overall attractiveness judgments (Field 2005).

RESULTS

Judges demonstrated concordance in their ratings (face: $W = .480$; body: $W = .441$; soundtrack: $W = .339$; all $p < .001$).

Linear regression was used to analyse the relationship between the ratings of the complete video and the various components (dependent variable: mean attractiveness score for complete video; independent variables: mean attractiveness scores for face, body and speech). The overall model was significant (adjusted $r^2 = .508$, $F = 17.85$, df = 49, $p < .001$). The face (standardised $\beta = .349$, $t = 3.07$, $p = .004$), body (standardised $\beta = .343$, $t = 2.89$, $p = .006$) and soundtrack (standardised $\beta = .278$, $t = 2.60$, $p = .013$) were independently and positively related to the rating of the complete video. Results are qualitatively identical if ratings of the neutral sentence replace ratings of the soundtrack. If target age is additionally included in the analysis, results are qualitatively identical, with the addition that age is independently and negatively related to the rating of the complete video (standardised $\beta$...
However, age is not significant \( (p > .1) \) with the exclusion of two cases (both males, aged 29 and 39) which appeared in this second analysis to have a disproportionate influence on the results (Mahalanobis distance > 14; Barnett and Lewis 1978).

Even though ratings were made by separate groups, there were significant correlations between the rating of the complete video and the ratings of the various components, both for all stimuli or for male and female stimuli separately (Table 1). Correction for multiple comparisons is not made because the correlations are not entirely independent of each other (i.e. if there is a correlation between face and video when all stimuli are judged, we would anticipate a similar relationship when the stimuli are restricted to the men or the women) and so a straightforward correction is not possible. Ratings of the various components were also positively and significantly correlated with each other (face/body: \( \tau = .350, p < .001 \); face/soundtrack: \( \tau = .196, p = .050 \); body/soundtrack: \( \tau = .197, p = .049 \)).

Table 1. Correlations of attractiveness ratings between individual components and rating of complete video

<table>
<thead>
<tr>
<th></th>
<th>all stimuli</th>
<th>male stimuli</th>
<th>female stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \tau )</td>
<td>( P )</td>
<td>( \tau )</td>
</tr>
<tr>
<td>face</td>
<td>.426</td>
<td>&lt; .001</td>
<td>.350</td>
</tr>
<tr>
<td>body</td>
<td>.460</td>
<td>&lt; .001</td>
<td>.425</td>
</tr>
<tr>
<td>soundtrack</td>
<td>.320</td>
<td>.001</td>
<td>.367</td>
</tr>
</tbody>
</table>

Next, we split the targets by sex, and re-ran the analyses. The overall model was significant (men/women: adjusted \( r^2 = .536/.714 \), \( F = 9.85/20.18 \), df = 23/23, \( p < .001/.001 \)). The face (men/women: standardised \( \beta = .463/.396 \), \( t = 2.92/2.76 \), \( p = .008/.012 \)), body (men/women standardised \( \beta = .453/.310 \), \( t = 2.96/2.14 \), \( p = .008/.045 \)) and, in women but not men, the soundtrack (men/women standardised \( \beta = .070/.415 \), \( t = .439/3.52 \), \( p = .666/.002 \)) were independently and positively related to the rating of the complete video. Here one man and one woman were excluded because analysis indicated that their cases may have had a disproportionate influence on results (Mahalanobis distance > 12; standardised DFBeta values > 1), although if they are retained in the model, results are qualitatively identical with the exception that ratings of men’s faces and women’s bodies are not significant predictors of ratings of the complete video. If target age is additionally included in the analysis, results are qualitatively identical; age is not a significant predictor (both \( p > .13 \)). Again, ratings of the various components were positively correlated with each other, although most values were non-significant (men: face/body: \( \tau = .184 \), \( p = .205 \); face/soundtrack: \( \tau = .222, p = .127 \); body/soundtrack: \( \tau = .260, p = .077 \);
women: face/body: $\tau = .464, p = .001$; face/soundtrack: $\tau = .158, p = .280$; body/soundtrack: $\tau = .149, p = .312$).

**DISCUSSION**

The current study builds upon previous research which has examined how different components such as the face and body combine to form a more holistic impression of personal attractiveness (e.g. Riggio et al. 1991; Peters, Rhodes and Simmons 2007). It also seeks to investigate conflicting findings showing that trait attractiveness in different modalities is (e.g. Brown, Cash and Noles 1986; Rikowski and Grammer 1999; Thornhill and Grammer 1999; Feinberg et al. 2005; Havlicek, Roberts and Flegr 2005; Saxton, Caryl and Roberts 2006; Feinberg et al. 2008; Roberts et al. in press) or is not (e.g. Raines, Hechtman and Rosenthal 1990a; Riggio et al. 1991; Oguchi and Kikuchi 1997; Rubenstein 2005; Lander 2008) positively correlated. We set out to categorise the relative importance of face, body and speech to overall judgments of attractiveness, and to investigate the relationships between attractiveness judgments of face, body and speech.

We recorded a set of target individuals as they introduced themselves on video. The self-introduction video was designed to correspond to zero acquaintance social contexts. We asked one group of raters to rate each of these complete video introductions for attractiveness. We asked a second, independent group of raters to rate the various components of the videos (a facial photograph of a neutral expression, masked to conceal the hair; the upper body; and the soundtrack of that video) for attractiveness.

Although one group of individuals rated the full videos, and a second group rated the face, body and speech, the mean ratings of the face, body and speech (in order of relative magnitude) were all significant predictors of the mean ratings of the complete videos. We also found that ratings of the individual components (face, body, speech) correlated significantly with each other. The correlation between the face and body demonstrated a medium-sized effect, while correlations with the speech were of a small effect size. There was some evidence that age related negatively to attractiveness ratings of the complete video. Rated female physical attractiveness tends to decline with age during adulthood (Symons 1995). In men, although age may show a positive relationship with attractive traits such as dominance (Keating 1985; Sadalla, Kenrick and Vershure 1987) and resources (Waynforth and Dunbar 1995), these traits may be less apparent from a short video clip. Age itself did not mediate the relationship between the attractiveness ratings of the components and the attractiveness ratings of the complete video: face, body and speech remained significant positive predictors when age was entered into the model.

When male and female targets were separated, results were broadly similar, with the exception that there was no evidence for an age effect, and in men, ratings
of the soundtrack were not significantly related to ratings of the complete video. Split by sex, correlations between ratings of the individual components (face, body, speech) tended to be positive but non-significant. This may be due to the smaller sample size, reducing power to detect real effects. Some studies have reported relationships in the attractiveness of physical traits in women but not in men (THORNHILL and GRAMMER 1999; PETERS, RHODES and SIMMONS 2007; LANDER 2008; PENTON-VOAK and CHANG 2008). Arguably, a woman’s mate value may be closely linked to the single dimension of fertility, whereas a man’s mate value may depend on a number of less tightly-knit dimensions (e.g. body size, testosterone levels, resources, parenting abilities). Accordingly, we might expect greater concordance in the attractiveness of different traits in women compared with men (see LANDER 2008; PENTON-VOAK and CHANG 2008), although this pattern was not apparent from our data. Future work might continue to investigate systematic sources of variability (e.g. the role of “multiple messages”; see MÖLLER and PO-MIANKOWSKI 1993; JOHNSTONE 1996; CANDOLIN 2003) between the attractiveness of different traits. For instance, in men, facial and bodily attractiveness, while linked, may also play different roles in indexing physical fitness (HÖNEKOPP et al. 2007).

Our results are not wholly consistent with previous research showing little relationship between attractiveness judged from a voice recording compared with attractiveness judged from a video recording (RAINES, HECHTMAN and ROSENTHAL 1990a, using videos complete with sound; also LANDER 2008 in same-sex judgments of silent videos), and with findings of low or no relationship between the attractiveness ratings of facial photographs and the attractiveness ratings of silent video recordings of the speaking faces (RUBENSTEIN 2005; LANDER 2008 and PENTON-VOAK and CHANG 2008 in recordings of males), but there are a number of methodological differences. Unlike the present study, vocal cues were only available to the raters of the videos in one of these previous studies (RAINES, HECHTMAN and ROSENTHAL 1990a). Vocal cues can modulate judgments of attractiveness (ZUCKERMAN, MIYAKE and HODGINS 1991; MIYAKE and ZUCKERMAN 1993; ZUCKERMAN, HODGINS and MIYAKE 1993). Previous studies also used shorter video extracts (between two and 10 s compared to the 20 s of the current study), and all but one (PENTON-VOAK and CHANG 2008) of the studies attempted to eliminate (RUBENSTEIN 2005; LANDER 2008) or pre-specify (RAINES, HECHTMAN and ROSENTHAL 1990a) emotional information and personal expression, qualities that affect perceptions of attractiveness (RAINES, HECHTMAN and ROSENTHAL 1990a, b; PENTON-VOAK and CHANG 2008). Exaggerated female facial movements are associated with enhanced female attractiveness (MORRISON et al. 2007; PENTON-VOAK and CHANG 2008), and it is possible that dynamic emotional or sexually dimorphic expression strengthens the relationship between static and dynamic attractiveness. In the present study, the inclusion of vocal information in longer and more naturalistic recordings where personal and emotional expression are allowed represents more realistically how an individual appears in the real world than do emotion-free or
soundless dynamic facial images. Although the current study makes use of videos, which are still a step removed from the real world, the finding that individual components such as static facial images relate to overall attractiveness suggests that ratings of individual components may yet be valid indices of overall individual attractiveness, and supports their use.

Future work may consider how attributes of the raters influence the relationship between overall personal attractiveness and the rated attractiveness of individual components such as a static facial photograph. While raters tend to agree on the attractiveness of others (LANGLOIS et al. 2000), this broad agreement can hide systematic individual variation. For instance, in comparing children, teenagers and young adults, the three groups give different weightings to facial and vocal attractiveness (ZUCKERMAN and HODGINS 1993), and only the older groups rate a man’s attractiveness similarly irrespective of whether they rate his face or voice (SAXTON, CARYL and ROBERTS 2006). In the same vein, only ratings from women at the high-fertility phase of the menstrual cycle gave rise to positive relationships between men’s body odour and visually-assessed attractiveness, and between men’s body odour and body symmetry (RIKOWSKI and GRAMMER 1999). The sex of the rater, and whether the rater is assessing same-sex or opposite-sex images, can also influence ratings (ROBERTS et al. in press). Cultural differences that affect ratings of attractiveness and that are apparent across modalities such as apparel and accent may increase the effect size of the correlation between individual components and holistic attractiveness in culturally heterogeneous stimuli, and vice versa. Our targets were culturally relatively homogeneous, since they were all recruited within one university. The study did not attempt to separate out more biological factors such as body shape or voice quality from more culturally-influenced factors such as choice of clothing or speech, and although this meant that raters in both conditions were privy to identical information, future research might seek to examine the impact of these various possible influences. We note though the lack of impact of the additional semantic and personal information available in the soundtrack, as demonstrated by the significant, large-effect correlations between ratings of the speech and ratings of the standardised sentence, and also by the consistency of our results irrespective of whether speech or standardised sentence ratings were used in the main regression analysis. Finally, although our video stimuli were designed to emulate a first meeting, other modalities may also impact upon first impressions, including bodily odour (e.g. Li et al. 2007; HAVLICEK and SAXTON in press) and bodily movement (BROWN et al. 2005).

Our results suggest that studies that employ visual stimuli alone to assess individual attractiveness in a mate choice context may be prone to a systematic source of experimental noise in that they do not consider the role of vocal attractiveness. Assessment of both vocal and visual components (c.f. ZUCKERMAN, MIYAKE and HODGINS 1991; MIYAKE and ZUCKERMAN 1993) would assist in a fuller understanding of an individual’s attractiveness. Nevertheless, in replicating results showing that the face is more important than the body (PETERS, RHODES and SIMMONS...
2007), and in extending this to show that face also has greater relative import than speech to ratings of personal attractiveness at zero acquaintance, our findings support the prevalent use of facial photographs as proxy measures of an individual’s attractiveness within human attraction and mate choice research. We recommend further research into the individual differences and the contextual effects that may moderate this relationship.

ACKNOWLEDGEMENTS

We thank M. Lyons and M. Guest for help with participant recruitment and session organisation and D.I. Perrett and B.P. Tiddeman for use of the Psychomorph program. TKS was supported by the University of Liverpool, is supported by an ESRC postdoctoral research fellowship, and thanks the Owen F. Aldis Scholarship Fund for its generous sponsorship of the research. AKM is supported by the BBSRC and Unilever R&D, Port Sunlight. HMR is supported by NERC.

REFERENCES


JEP 7(2009)1


