

# Implicitly Perceived Vocal Attractiveness Modulates Prefrontal Cortex Activity

Patricia E. G. Bestelmeyer<sup>1</sup>, Marianne Latinus<sup>1</sup>, Laetitia Bruckert<sup>2</sup>, Julien Rouger<sup>3</sup>, Frances Crabbe<sup>1</sup> and Pascal Belin<sup>1,4</sup>

<sup>1</sup>Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK, <sup>2</sup>Sciences de l'Homme et de la Société, Université Pierre-Mendès-France, 38400 Grenoble, France, <sup>3</sup>Department of Cognitive Neuroscience, University of Maastricht, 6200 MD Maastricht, the Netherlands and <sup>4</sup>International Laboratories for Brain, Music and Sound Research, Université de Montréal and McGill University, Montréal H3C 3J7, Canada

Address correspondence to Dr Patricia E. G. Bestelmeyer, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK. Email: patricia.bestelmeyer@glasgow.ac.uk; new address: School of Psychology, Bangor University, Bangor, LL57 2AS, UK

**Social interactions involve more than “just” language. As important is a more primitive nonlinguistic mode of communication acting in parallel with linguistic processes and driving our decisions to a much higher degree than is generally suspected. Amongst the “honest signals” that influence our behavior is perceived vocal attractiveness. Not only does vocal attractiveness reflect important biological characteristics of the speaker, it also influences our social perceptions according to the “what sounds beautiful is good” phenomenon. Despite the widespread influence of vocal attractiveness on social interactions revealed by behavioral studies, its neural underpinnings are yet unknown. We measured brain activity while participants listened to a series of vocal sounds (“ah”) and performed an unrelated task. We found that voice-sensitive auditory and inferior frontal regions were strongly correlated with implicitly perceived vocal attractiveness. While the involvement of auditory areas reflected the processing of acoustic contributors to vocal attractiveness (“distance to mean” and spectrotemporal regularity), activity in inferior prefrontal regions (traditionally involved in speech processes) reflected the overall perceived attractiveness of the voices despite their lack of linguistic content. These results suggest the strong influence of hidden nonlinguistic aspects of communication signals on cerebral activity and provide an objective measure of this influence.**

**Keywords:** auditory cortex, Broca's area, implicit, nonverbal communication, voice

## Introduction

In addition to linguistic content, voices convey other kinds of socially relevant information that appear to be processed by humans as part of a more primitive nonlinguistic mode of communication (Buchanan 2009). In particular, vocal attractiveness is one such example of simple hidden social signals that have a tremendous, if mostly unconscious influence on social interactions. Vocal attractiveness appears to be a phenotypic marker of health and reproductive fitness (Hughes et al. 2002, 2004) and is linked with increased numbers of sexual partners, decreased reported age of first sexual contact, and a larger number of extramarital affairs (Hughes et al. 2004). Vocal attractiveness ratings are highly reliable across participants (Zuckerman and Driver 1989), and individuals with attractive voices are credited with socially more favorable traits, such as likeability, competence, and conscientiousness (Zuckerman and Driver 1989; Zuckerman et al. 1990). This phenomenon is known as the “what sounds beautiful is good” stereotype (Zuckerman and Driver 1989).

Despite the wide-ranging influence of vocal attractiveness on our social interactions, the cerebral underpinnings are unknown. The main goal of the present study was to examine the extent to which implicitly perceived vocal attractiveness would modulate cerebral activity. We used functional magnetic resonance imaging (fMRI) to measure participants' cerebral oxygenation level while they listened to vocal sounds (“ah”) produced by different voices ( $n = 64$ ) and performed an irrelevant tone detection task. These voices were independently rated for attractiveness. We predicted that the behaviorally robust effects of attractiveness would translate into reliable neural differences associated with implicitly perceived vocal attractiveness. We expected these differences to particularly involve the voice-sensitive “temporal voice areas” (TVA) of auditory cortex (Belin et al. 2000; Grandjean et al. 2005; Ethofer et al. 2009). Since vocal attractiveness has never been investigated before we made no specific predictions as to which higher level auditory network would be involved. However, a previous study on the perception of socially significant vocal sounds (vocal affect) suggests that bilateral inferior prefrontal cortex is involved in the processing of affective versus neutral prosody (Kotz et al. 2003). In addition, Schirmer and Kotz (2006) have argued that after an initial low-level processing in auditory cortex, inferior frontal regions are involved in vocal prosody perception, and most recently, a study on the connectivity between auditory cortex and inferior prefrontal areas has demonstrated that vocal emotion compared with neutral prosody enhanced the functional coupling of high-level auditory cortex with ipsilateral inferior frontal gyrus (IFG; Ethofer et al. forthcoming). We therefore expected that bilateral IFG could be a possible candidate for the processing of other socially significant cues, such as vocal attractiveness.

We have recently examined behavioral gender-independent mechanisms of vocal beauty (Bruckert et al. 2010). Morphed vocal composites or averages generated from increasing numbers of voices were judged to be more attractive than the original voices they were constructed from. We identified 2 largely independent acoustic contributors to the attractiveness ratings: the voice's “distance to mean” reflecting its degree of similarity with the average voice and the voice's spectrotemporal regularity (measured by its harmonics-to-noise ratio [HNR] in dB). Voices more similar to the average voice as well as more regular “smoother” voice-textures were perceived as more attractive (Bruckert et al. 2010). A secondary goal of the study was therefore to examine the contribution of these acoustic parameters on the pattern of cerebral response. We predicted that these 2 independent acoustic contributors to vocal attractiveness would engage neurally dissociable regions of auditory cortex.

## Materials and Methods

### Participants

Twenty volunteers from the under- and postgraduate community of the University of Glasgow took part (11 females, mean age = 23.15, range = 18–31 years, standard deviation [SD] = 3.75). Participants reported normal hearing and were reimbursed £12 for their time (£6/hour). Informed consent was obtained from all individuals, and the study protocol was approved by the local ethics committee.

### Stimuli and fMRI Paradigm

Voice stimuli were the ones used in Bruckert et al. (2010). They consisted of digital samples (16 bit, mono, 16 kHz sampling rate) of 32 male and 32 female adult speakers producing the sound “had” (Hillenbrand et al. 1995). To avoid averaging artifacts, we used Adobe Audition to remove the release burst in the final “d.” Stimuli were then normalized for acoustic power root mean square and ranged between 201 and 477 ms in duration (mean = 311 ms; SD = 50). Voice averaging was performed using STRAIGHT (Kawahara and Matsui 2003) in MatlabR2007 (The MathWorks Inc.). STRAIGHT performs an instantaneous pitch-adaptive spectral smoothing of each stimulus for separation of contributions to the voice signal arising from the glottal source (including f0) versus supralaryngeal filtering (distribution of spectral peaks, including the first formant, F1). Voice stimuli were decomposed by STRAIGHT into 5 parameters (f0, frequency, duration, spectrotemporal density, and aperiodicity) that can be manipulated independently of one another. For each stimulus, we manually identified landmarks in the time-frequency domain of the sounds and located these in correspondence with each other across all voices. Morphed stimuli were then generated by resynthesis based on the interpolation (linear for time; logarithmic for f0, frequency, and amplitude) of these time-frequency landmark templates. For each speaker sex, 32 individual voices were randomly paired with one another to generate sixteen 2-voice composites. This process was repeated at subsequent degrees of averaging to yield eight 4-voice composites, four 8-voice composites, two 16-voice composites, and a single 32-voice composite for each gender (see Bruckert et al. (2010) for further details).

The fMRI paradigm consisted of 4 blocks of 86 trials (2 blocks per gender); 63 voices of one gender (32 natural and 31 composites voices), 7 pure tones (1000 Hz), and 9 additional silences were randomly presented binaurally using the MRI compatible NNL headphone system (NordicNeuroLab, Inc.) at 80 dB sound pressure level (C). Participants were asked to keep their eyes closed, listen passively to the voices, and press a button every time they heard a pure tone. Pure tone detection performance during the scanning sessions was at 91.6% (SD = 9.92). After the completion of the fMRI paradigm, participants rated the voices for attractiveness on an analog scale ranging from “extremely unattractive” to “extremely attractive.”

For the parametric modulation analysis of attractiveness, we used a set of independent ratings obtained in better listening conditions and without prior exposure to the stimuli (Bruckert et al. 2010). Participants recruited by Bruckert et al. (2010) were of the same age group and had the same gender balance as participants recruited for the present experiment. The interrater reliability of these ratings was high (Cronbach’s alpha of 0.98; Bruckert et al. 2010). Attractiveness judgments of each individual rater were converted to z-scores using the average and SD of the ratings for the 64 natural unaveraged male and female voices. Attractiveness z-scores were then averaged across raters and used as a regressor in the fMRI analyses.

Acoustic voice similarity can be expressed as the Euclidian distance within a logarithmic 2D “voice space” with axes defined by the fundamental frequency, f0 (perceived as voice pitch, reflecting laryngeal excitation) against the first formant frequency, F1 (related to vocal tract filtering and perceived as voice timbre). In this voice space, voices that are closer together are perceived as more similar, as indicated by multidimensional scaling analyses of voice identity dissimilarity ratings (Baumann and Belin 2010). We computed the distance in that voice space between each voice and the same-gender prototype (the 32-voice composite) and used these values to find the relationship between blood oxygen level-dependent (BOLD) signal and

distance to the vocal prototype. In addition to attractiveness ratings and distance measures, we used a measure of voice “regularity,” HNR (measured in decibel), as another regressor. HNR measures the energy of the periodic component of the voice relative to the energy of its noisy, or aperiodic, part, using an autocorrelation algorithm. Measures of f0, F1, and HNR were performed using Praat (Boersma and Weenink 2005).

### Image Acquisition and Analysis

All scans were acquired in a 3.0T Siemens Tim Trio scanner using a 12-channel head coil. Whole-brain  $T_1$ -weighted anatomical scans were performed using fast gradient echo known as  $T_1$  “magnetization prepared rapid gradient echo” consisting of 192 axial slices of 1 mm thickness with an inplane resolution of  $1 \times 1 \times 1$  (field of view [FOV] = 256) and a matrix of  $256 \times 256$  performed at the end of the experimental session.  $T_2$ -weighted functional scans using an echoplanar imaging sequence were acquired using an interleaved ascending sequence consisting of 32 slices of 3 mm thickness (0.3 mm gap) with an inplane resolution of  $3 \times 3 \times 3$  (FOV = 210) and a matrix of  $70 \times 70$ . The 4 runs of the fast event-related experimental scan (repetition time [TR] = 2 s, acquisition time [TA] = 1.5 s, echo time [TE] = 30 ms) consisted of 180 volumes each while the voice localizer (TR = 2 s; TE = 30 ms) scan (block design) consisted of 310 volumes and allows reliable identification of the TVA using the vocal versus nonvocal contrast (see Belin et al. (2000) for details and <http://vnl.psy.gla.ac.uk> for download). In the experimental scan, stimuli were presented in the middle of each second silent gap of 500 ms while every other silent gap consisted of an additional silence (see Fig. 1 for illustration). For the voice localizer, the sounds were superimposed on the scanner noise.

All MRI data were analyzed using SPM5 (Wellcome Department of Imaging Neuroscience, 1994–2007). Preprocessing of functional scans consisted of corrections for head motion (spatial realignment; trilinear interpolation), and scans were realigned to the first volume of the last session (i.e., the volume closest to the anatomical scan). Functional runs were then coregistered to their corresponding individual anatomical scans. Functional (3 mm isotropic voxels) and anatomical (1 mm isotropic voxels) data were transformed to Montreal Neurological Institute (MNI) space after segmentation of the anatomical scans. Normalized data were spatially smoothed by applying a Gaussian kernel of 8 mm full-width at half-maximum.

For the parametric modulation analyses, we linearly rescaled the variables of interest so that all measurements were on a comparable scale lying between 0 and 1 using the “Min-Max” normalization. This transformation preserves the relationships among the original data values and was essential for the last 3 models to ensure that the regressors have equivalent weights. In a parametric modulation analysis, additional regressors are orthogonalized from left to right in a given matrix so that if a weight of 1 is attributed to the regressor of interest (e.g., attractiveness), the variance of prior regressors (e.g., voice duration or HNR) is removed (Büchel et al. 1998). Voice duration correlated positively with BOLD signal. It was therefore treated as a confound and entered as a parametric modulator before any variables of interest so that any variance explained by voice duration was removed. However, the results do not differ significantly when sound duration is not regressed out.

The first 3 models assessed linear relationships between neural activity and 1) attractiveness ratings, 2) distance to mean, and 3) HNR, respectively. Here, the matrix contained the onsets of all voices as the first regressor. The following 2 regressors were entered as parametric modulators: Voice duration followed by either attractiveness ratings or distance measures or HNR (depending on the model of interest). Moreover, using a similar matrix, we computed a model assessing any



**Figure 1.** Diagram depicting the timeline of stimulus presentation.

quadratic relationships between attractiveness ratings and BOLD signal to be consistent with studies in the face literature (Winston et al. 2007).

The remaining 3 models examined linear relationships between BOLD signal and attractiveness ratings in a model additionally containing either 1) distance to mean, 2) HNR regressors, or 3) both (compared with previous models in which we analyzed each parametric modulator on its own in a single matrix). Thus, the matrix of these models contained the same first and second regressors as previous models plus regressors representing one or both acoustic measures followed by the attractiveness regressor.

All models contained the pure tone onsets and 6 movement parameters as the last 7 regressors. Reported results are from whole-brain analyses. Percent signal change was calculated in corresponding maxima within a sphere (radius of 6 mm) using RFXplot (Gläscher 2009). Results are reported descriptively at a height threshold of  $P < 0.001$  (uncorrected) with an extent threshold of more than 15 voxels unless stated otherwise. Results are illustrated on a single-subject anatomical scan using SPM5 (Wellcome Department of Imaging Neuroscience, 1994–2007). Statistical significance was assessed at the cluster level with a height threshold of  $P < 0.05$  (corresponding to a cluster size of at least 50 voxels), FWE-corrected for multiple comparisons across the whole brain. We employed a small volume correction for clusters in bilateral IFG and corrected for multiple comparisons within the IFG using a digital atlas of the human brain (Automated Anatomical Labeling; Tzourio-Mazoyer et al. 2002). Positive correlations are plotted in red while negative correlations are plotted in blue. Bilateral TVA are shown in yellow. Illustrations are depicted in neurological convention.

## Results

### *Parametric Modulation Analyses for Attractiveness Ratings, Distance to Mean, and HNR*

We found a significant correlation between implicitly perceived attractiveness and a well-defined network of cortical regions. A significant negative correlation between attractiveness ratings and BOLD response was observed in bilateral superior temporal sulci/gyri (STS/STG) as well as in IFG of the right hemisphere, that is, these regions were more responsive to increasingly unattractive voices. Plots of percent signal change as a function of attractiveness illustrate this negative correlation in the right IFG (Fig. 2A) and in areas overlapping with bilateral TVA (green shows the overlap between TVA and the activity correlated with attractiveness ratings). Lowering the threshold to  $P < 0.005$  (uncorrected) revealed a positive correlation between BOLD signal and attractiveness ratings in bilateral fusiform gyri (FG) and left midoccipital cortex (Fig. 2B). Coordinates of the maximum activity in each cluster of the FG were close to, but different from, the area defined as the fusiform face area (FFA) in other studies (MNI coordinates for peaks in the current study: FG right: 27, -39, -12; FG left: -30, -42, -18 compared with MNI coordinates of clusters identified as FFA by Summerfield et al. (2008), for example: FFA right: 46, -50, -24, FFA left: -44, -44, -20). There were no significant quadratic correlations for the attractiveness modulator.

Bruckert et al. (2010) have shown that part of the variance of attractiveness ratings can be explained by 2 acoustic properties: distance to mean and HNR. While distance to mean explained more of the variance, both measures contributed significantly (distance to mean:  $R^2 = 0.36$ ,  $F_{1,125} = 69.74$ ,  $P < 0.0001$ ; HNR:  $R^2 = 0.13$ ,  $F_{1,125} = 18.94$ ,  $P < 0.0001$ ). Taken together, distance to mean (Fig. 3A) and HNR (Fig. 3B) explained 40% of the variance in attractiveness ratings ( $R^2 = 0.40$ ,  $F_{2,125} = 41.76$ ,  $P < 0.0001$ ).

We computed 2 additional models to investigate the relationship between these 2 acoustic properties and BOLD

activity. The results of the parametric modulation analysis with distance to mean as a third regressor were in line with the attractiveness finding. It revealed a significant positive correlation in bilateral STS/STG, that is, voices which were further away from the vocal prototype in the voice space (and therefore less attractive) elicited a greater response in bilateral STS/STG (illustrated in Fig. 3C). Similarly, voices with a higher HNR (i.e., more regular texture) were related to a decrease in BOLD signal in bilateral STS/STG (Fig. 3D; although this negative correlation was marginally significant only in the right TVA after FWE-correction;  $P = 0.059$ ). The activity correlated with distance to mean and HNR overlapped with the bilateral TVA. However, activity correlated with HNR was weaker and located more anterior than the areas responsive to distance to mean. Supplementary Table 1 summarizes cluster size and location for the attractiveness, distance to mean, and HNR models.

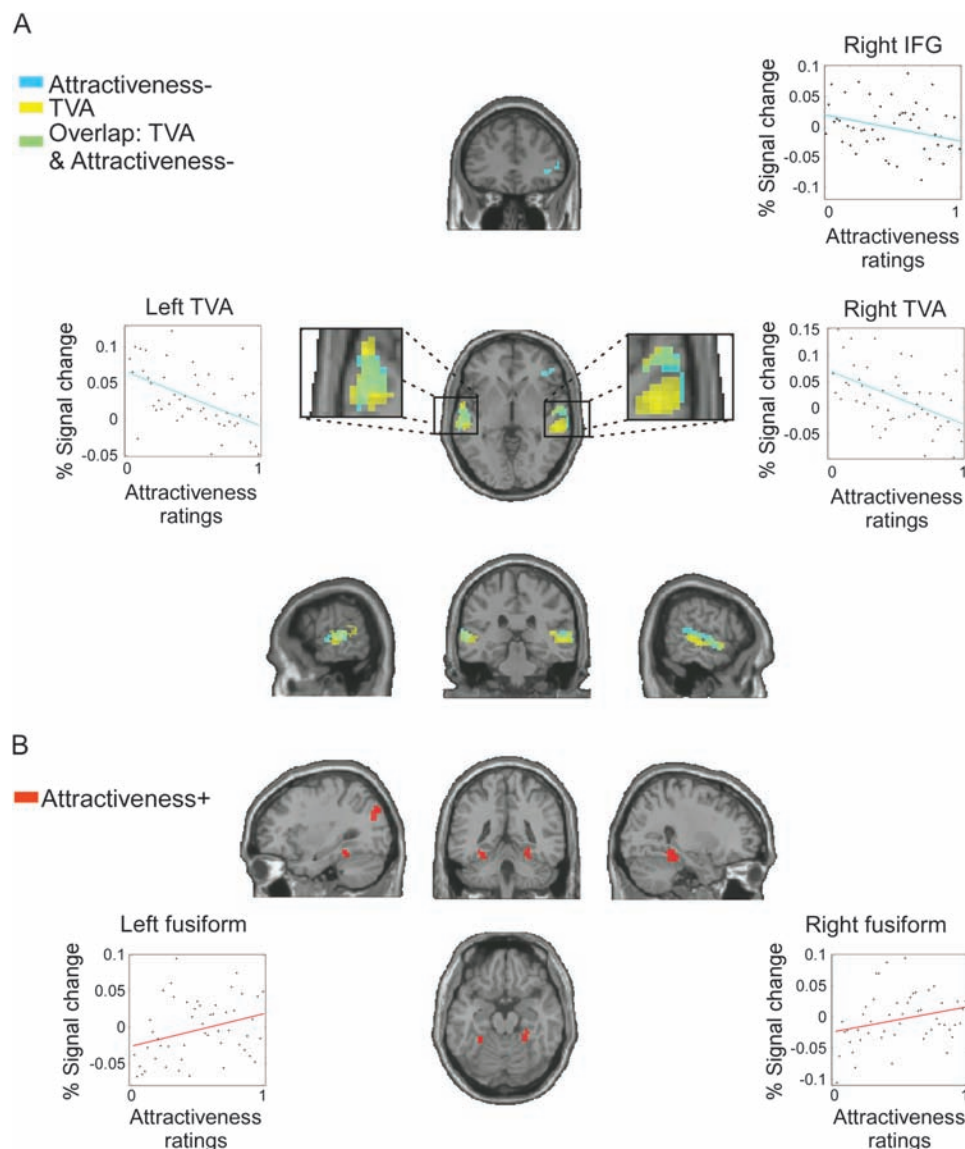
### *Parametric Modulation Analysis for Attractiveness after Removing Variance Explained by One or Both Objective Acoustic Measures*

Distance to mean and HNR are largely independent dimensions in the natural voices (Bruckert et al. 2010). However, due to the averaging process, we used to generate increasingly attractive voices, the 2 acoustic factors are no longer orthogonal when considering all voices together (i.e., the natural and averaged voices). In order to disentangle the contribution of these 2 acoustic properties, we performed 3 additional parametric modulation analyses by varying the order of the parametric modulators (with orthogonalization of each regressor relative to the previous ones; see Materials and Methods). In other words, we explored the relation between BOLD signal as a function of attractiveness ratings after the contribution of distance to mean, HNR, or both factors had been regressed out. Some activity remained in bilateral TVA after variance explained by either distance to mean or HNR had been removed. Strikingly, activity in the right IFG, particularly its triangular part, remained when either distance to mean or HNR or both factors had been explained. After both acoustic contributors to vocal attractiveness had been removed the areas, which seem to be significantly correlated with the perceptual experience of attractiveness were bilateral IFG (Fig. 3E). Supplementary Table 2 summarizes cluster size and location for the 3 models.

## Discussion

We were interested in the influence of implicitly perceived vocal attractiveness on the brain. We found that despite no linguistic content and the implicit nature of the task (subjects were asked to detect an occasional pure tone stimulus) perceived vocal attractiveness significantly correlated with activity in a well-circumscribed network of areas including higher level auditory cortex and inferior prefrontal regions. Correlations in the IFG remained significant when the known acoustic contributors to vocal attractiveness were taken into account. This finding suggests that these regions are involved in a perceptual representation of vocal attractiveness relatively abstracted from its low-level acoustic determinants. These results imply the strong influence of vocal attractiveness on cerebral activity, even in implicit conditions, and provide an objective measure for this influence.





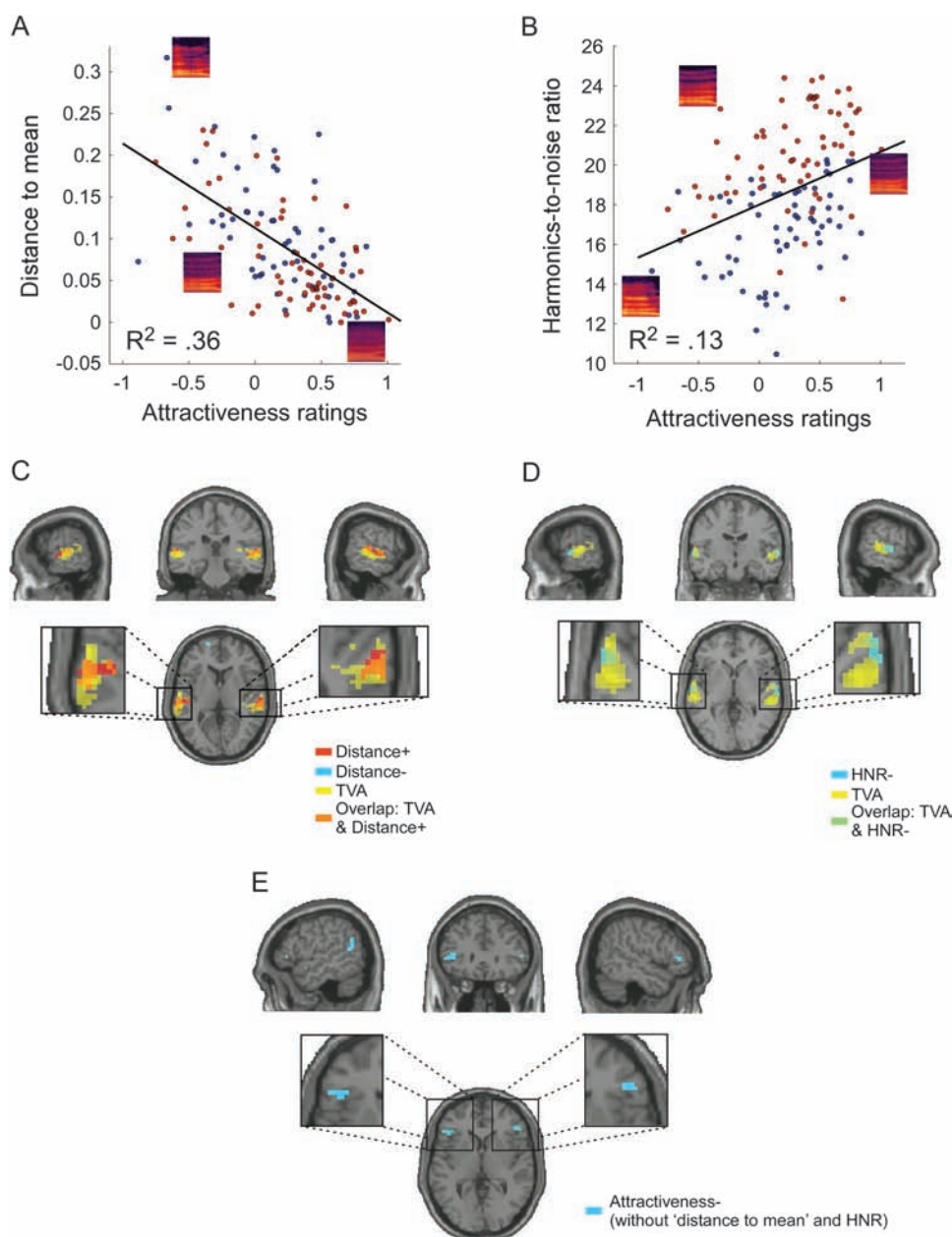
**Figure 2.** (A) Negative correlations between attractiveness ratings and BOLD response (blue) in right IFG as well as bilateral STS/STG. The overlap between negative correlations and bilateral TVA (yellow) is shown in green. Scatterplots illustrate average attractiveness ratings of the voices (represented as dots) as a function of percent signal change (MNI coordinates of sections:  $\pm 61, -18, 1$ ). (B) Positive correlations between attractiveness ratings and BOLD response (red) in bilateral fusiform areas and left midoccipital regions at a more lenient threshold ( $P < 0.005$ , uncorr). Scatterplots illustrate average attractiveness ratings as a function of percent signal change (MNI coordinates of sections:  $\pm 28, -2, -17$ ).

### Correlation between Cerebral Response to Voices and Implicitly Perceived Vocal Attractiveness

Voices independently rated as less attractive correlated with greater activity in voice-sensitive areas (bilateral STS/STG) and right IFG triangularis. Voice-sensitive areas are known to respond more strongly to vocal sounds than to other sound categories (Belin et al. 2000; Lewis et al. 2009), while the IFG is part of Broca's area and is typically linked to speech and language processes (Broca 1861; Musso et al. 2003; Fadiga et al. 2009; Price and Drevets 2010). Conversely, at a slightly higher threshold, more attractive voices were related to increased activation in bilateral FG and left midoccipital regions. Bilateral FG tend to be associated with processes in the visual modality, predominately with the processing of faces (e.g., Kanwisher et al. 1997). Although the activation we report in the FG does not overlap "exactly" and is slightly more medial than the area

many studies define as the FFA, it is nevertheless surprising that our study focusing on "vocal" attractiveness resulted in activation of visual areas.

Previous results on the neural substrates of "facial" attractiveness are often inconsistent. Only some studies find enhanced responses to facial attractiveness in face-sensitive areas, such as the FFA (Iaria et al. 2008; Chatterjee et al. 2009). Many studies also find increased responses to unattractive faces (e.g., O'Doherty et al. 2003). The discrepant findings in the face literature may be explained by differences in stimulus set, design, and task instructions. Our finding of increased response to decreasingly attractive voices in the right IFG is in line with O'Doherty et al.'s (2003) investigation into implicit perceptions of facial attractiveness. Although tentative and requiring replication in a study assessing both facial and vocal beauty using similar setup and design, the fMRI patterns that emerged in



**Figure 3.** (A) Negative correlations between behavioral attractiveness ratings and measures of distance to mean for male (blue) and female (red) voices. (B) Positive correlations between behavioral attractiveness ratings and spectrotemporal regularity (as measured by HNR; dB) for male (blue) and female (red) voices. (C) Positive correlation between measures of distance to mean and BOLD response (red) in bilateral STS/STG. The overlap between bilateral TVA (yellow) and the areas responsive to modulations in distance to mean is shown in orange (MNI coordinates of sections:  $\pm 63$ ,  $-27$ ,  $4$ ). (D) Negative correlations (blue) between measures of spectrotemporal regularity and BOLD response in bilateral STS/STG. The overlap between bilateral TVA (yellow) and the areas responsive to modulations in HNR (dB) is shown in green (MNI coordinates of sections:  $\pm 63$ ,  $-12$ ,  $-1$ ). (E) Negative correlations between attractiveness ratings and BOLD response (blue) in bilateral IFG and left angular gyrus after variance explained by distance to mean and spectrotemporal regularity have been removed.

this study and studies on facial attractiveness could suggest that at least the FG and right IFG are responsive to attractiveness supramodally and engage a broad neural network dealing with attractiveness. Regarding the activity in bilateral FG, there is at least one alternative explanation. Daily life typically requires voice and face processes to be engaged in parallel. Although anatomically separate, various face models also assume an interaction between face and voice modules (e.g., Burton et al. 1990; Ellis et al. 1997) with evidence for reciprocal functional connections between them (von Kriegstein et al. 2005). Given that facial and vocal attractiveness are correlated (Collins and

Missing 2003), it is possible that auditory input is coupled to or is translated into a visual representation (von Kriegstein et al. 2003, 2005).

#### Effects of Acoustic Contributors to Vocal Attractiveness

Bruckert et al. (2010) discovered 2 distinct acoustic factors, which contribute to the perception of vocal attractiveness: distance of a voice to the vocal average in the acoustic f0-F1 voice space, and voice spectrotemporal regularity as measured by HNR. We found neural correlates for the processing of both objective measures. Mirroring the attractiveness results, we

found positive correlations between BOLD signal and the distance to the average voice in bilateral STS/STG. In other words, the more distinct or different a voice is from the vocal prototype, the more unattractive it will be perceived, and the larger the activation in bilateral STS/STG. It is possible that, similar to faces (see Leopold et al. 2001, 2005; Löffler et al. 2005), voices are coded in terms of a prototype against which individuating information, such as attractiveness is coded. This is speculative and necessitates further more in depth research in which vocal distinctiveness is varied with greater control than in the current study.

Voices with more irregular texture (perceived as more unattractive) were related to increased activity in bilateral STS/STG. An important study by Lewis et al. (2009) also found harmonicity-sensitive regions in the auditory cortex. Here, increasing HNR in simple synthesized sounds and animal vocalizations increased activation in bilateral Heschl's gyri and medial STS including regions, which respond preferentially to human nonspeech vocalizations. Lewis et al. (2009) also report significant negative correlations at the single-subject level between HNR and BOLD signal, which were located along the medial wall of the lateral sulcus, however, these negatively correlated patterns were not significant once the data were group averaged and were therefore not discussed further. To our knowledge, the present study is the first to show that texture-sensitive regions are activated parametrically within the category of human nonspeech vocalizations. The reasons for opposite activation patterns for synthetic sounds or animal vocalizations reported in Lewis et al.'s (2009) study, and human nonspeech vocalizations shown in the present one are unclear and warrant further investigation.

Generally, activity in the STS/STG in response to attractiveness, distance to mean and HNR overlapped largely with voice-sensitive areas. Despite the fact that both acoustic properties were processed in bilateral TVA, it was possible to dissociate them and show that although correlated distinct neural areas are sensitive to them. HNR was processed more anterior than distance to mean in voice-sensitive regions. The greater and more significant activity in response to the distinctiveness of a given voice compared with its regularity reflects our behavioral findings in which distance to mean explained almost 3 times more of the variance of the attractiveness ratings than HNR. Overall, distinct neural signatures were evident for attractiveness, distance to mean, and spectrotemporal regularity of the voice.

### ***Prefrontal Cortex Modulation by Perceived Vocal Attractiveness***

When removing the variance explained by distance to mean and HNR, no activity remained in bilateral TVA suggesting that these areas mainly deal with the "lower level" acoustic vocal features contributing to the perception of attractiveness (sound duration was accounted for in all models). This result is similar to that reported by Wiethoff et al. (2008) who showed relationships between the processing of emotional intonation and responses in the superior temporal cortex. Analogous to the results described here, this relationship was abolished once acoustic parameters, such as pitch and sound duration, were removed. Both studies suggest that areas in the STS/STG may deal with low-level acoustics of vocal sounds. It is important to note that there may of course be other, so far unknown, acoustic contributors to vocal attractiveness than the ones we

have investigated here, but it is striking that all activity in the TVA is abolished when accounting for distance to mean (incorporating F0 and f1) and HNR.

Once we had removed the variance explained by acoustic parameters, significant activity remained in the triangular part of the bilateral IFG. This region is part of Broca's area (Anwander et al. 2007; Grodzinsky and Santi 2008) and is strongly connected to sensory cortex (Frey et al. 2008; Petrides and Pandya 2009). In addition to its involvement in language perception, bilateral activity in Broca's area has been linked to auditory working memory in which increased task demands correlate with increased activity (Martinkauppi et al. 2000; Arnott et al. 2005; Schulze et al. 2011). There is also evidence that implies Broca's area in serving integrating functions between multimodal sensory information (Price and Drevets 2010) and paralinguistic features, such as affective prosody (Ethofer et al. forthcoming; Kotz et al. 2003). Based on the aforementioned research, our results may suggest that increasingly unattractive voices demand larger processing resources and may point toward the role of the IFG triangularis as being involved not only in the processing of language and affective prosody but also in integrating acoustic information received from bilateral TVA into a unified percept of attractiveness. Importantly, the pattern of results emerging after removing the variance explained by the acoustic contributors to vocal attractiveness shows that not any deviation from the average voice can explain the pattern of activity in bilateral IFG.

### ***Limitations and Future Endeavors***

The current study is the first to investigate the cerebral correlates of the implicit perception of vocal attractiveness and its acoustic contributors. However, many open questions remain. For example, how do our findings relate to natural speech, how are vocal attractiveness and speech processed in the IFG during natural communication, is the IFG involved in the processing of other socially relevant information in addition to vocal affect and attractiveness? Further, the current study used previously acquired ratings of vocal attractiveness as parametric modulators in the analysis to avoid any spurious effects of familiarity on vocal attractiveness ratings. Although previous research has shown that interrater reliability of vocal attractiveness is very high (Bruckert et al. 2010), some interrater variability is likely. Another interesting question is therefore how individual variability in the perception of attractiveness modulates cerebral activity. An explicit judgment task of vocal attractiveness inside the scanner may be most suitable to address this question.

### ***Conclusion***

We show for the first time that vocal attractiveness modulates the activity of a cortical network comparable to that engaged consciously by speech perception, particularly bilateral inferior prefrontal regions. This modulation of prefrontal activity was observed despite the participant's lack of awareness of the dimensions being manipulated, illustrating the pervasive influence of implicitly perceived vocal attractiveness on cerebral activity. Further, the results identify different cerebral substrates for the processing of the acoustic contributors (distance to mean and spectrotemporal regularity) to vocal attractiveness. Our results provide objective



support for the existence of a primitive nonlinguistic mode of communication that is at the root of human social interaction. This research is not only important for the general advancement of knowledge and understanding about what neural network is engaged during vocal attractiveness perception but also has important industrial implications. We increasingly communicate with computers. In the emerging domain of automated social signal processing, our results are important for engineers who need to know what an attractive voice is and how best to automatically extract and process different types of vocal information, such as attractiveness, affect, or identity.

## Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

## Funding

Economic and Social Research Council/Medical Research Council grant (RES-060-25-0010)

## Notes

*Conflict of Interest:* None declared.

## References

- Anwander A, Tittgemeyer M, von Cramon DY, Friederici AD, Knosche TR. 2007. Connectivity-based parcellation of Broca's area. *Cereb Cortex*. 17:816–825.
- Arnott SR, Grady CL, Hevenor SJ, Graham S, Alain C. 2005. The functional organization of auditory working memory as revealed by fMRI. *J Cogn Neurosci*. 17:819–831.
- Baumann O, Belin P. 2010. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychol Res*. 74:110–120.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. *Nature*. 403:309–312.
- Boersma P, Weenink D. 2005. Praat: doing phonetics by computer [Computer program]. Version 5.0.35. Available from: <http://www.praat.org/>. Accessed October 2008.
- Broca P. 1861. Remarques sur le siege defaulte de langage articule suivies d'une observation d'aphemie (perte de la parole). *Bull Soc Anthropol*. 2:330–337.
- Bruckert L, Bestelmeyer P, Latinus M, Rouger J, Charest I, Rousselet GA, Kawahara H, Belin P. 2010. Vocal attractiveness increases by averaging. *Curr Biol*. 20:116–120.
- Buchanan M. 2009. Social signals. *Nature*. 457:528–530.
- Büchel C, Holmes AP, Rees G, Friston KJ. 1998. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*. 8:140–148.
- Burton AM, Bruce V, Johnston RA. 1990. Understanding face recognition with an interactive activation model. *Br J Psychol*. 81:361–380.
- Chatterjee A, Thomas A, Smith SE, Aguirre GK. 2009. The neural response to facial attractiveness. *Neuropsychologia*. 23:135–143.
- Collins S, Missing C. 2003. Vocal and visual attractiveness are related in women. *Anim Behav*. 6:997–1004.
- Ellis HD, Jones DM, Mosdell N. 1997. Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol*. 88:143–156.
- Ethofer T, Bertscher J, Gschwind M, Kreifelts B, Wildgruber D, Vuilleumier P. 2011. Emotional voice areas: anatomic location, functional properties and structural connections revealed by combined fMRI/DTI. 30. doi:10.1093/cercor/bhr113.
- Ethofer T, De Ville DV, Scherer K, Vuilleumier P. 2009. Decoding of emotional information in voice-sensitive cortices. *Curr Biol*. 19:1028–1033.
- Fadiga L, Craighero L, D'Ausilio A. 2009. Broca's area in language, action, and music. *Ann N Y Acad Sci*. 1169:448–458.
- Frey S, Campbell JSW, Pike GB, Petrides M. 2008. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci*. 28:11435–11444.
- Gläscher J. 2009. Visualization of group inference data in functional neuroimaging. *Neuroinformatics*. 7:73–82.
- Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci*. 8:145–146.
- Grodzinsky Y, Santi A. 2008. The battle for Broca's region. *Trends Cogn Sci*. 12:474–480.
- Hillenbrand JM, Getty LA, Clark MJ, Wheeler K. 1995. Acoustic characteristics of American English vowels. *J Acoust Soc Am*. 97:1300–1313.
- Hughes SM, Dispenza F, Gallup GGJ. 2004. Ratings of voice attractiveness predict sexual behavior and body configuration. *Evol Hum Behav*. 24:295–304.
- Hughes SM, Harrison MA, Gallup GGJ. 2002. The sound of symmetry: voice as a marker of developmental instability. *Evol Hum Behav*. 23:173–180.
- Iaria G, Fox CJ, Waite CT, Aharon I, Barton JJS. 2008. The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: neuropsychological and neuroimaging evidence. *Neuroscience*. 155:409–422.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.
- Kawahara H, Matsui H. 2003. Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*; Hong Kong. p. 256–259.
- Kotz S, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD. 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang*. 86:366–376.
- Leopold DA, O'Toole AJ, Vetter T, Blanz V. 2001. Prototype-referenced shape encoding revealed by high-level after effects. *Nat Neurosci*. 4:89–94.
- Leopold DA, Rhodes G, Muller KM, Jeffery L. 2005. The dynamics of visual adaptation to faces. *Proc R Soc Lond B Biol Sci*. 272: 897–904.
- Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA. 2009. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J Neurosci*. 29:2283–2296.
- Löffler G, Yourganov G, Wilkinson F, Wilson HR. 2005. fMRI evidence for the neural representation of faces. *Nat Neurosci*. 8:1386–1390.
- Martinkauppi S, Rama P, Aronen HJ, Korvenoja A, Carlson S. 2000. Working memory of auditory localization. *Cereb Cortex*. 10:889–898.
- Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Büchel C, Weiller C. 2003. Broca's area and the language instinct. *Nat Neurosci*. 6:774–781.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ. 2003. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*. 41:147–155.
- Petrides M, Pandya DN. 2009. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol*. 7:e1000170.
- Price JL, Drevets WC. 2010. Neurocircuitry of mood disorders. *Neuropsychopharmacology*. 35:192–216.
- Schirmer A, Kotz SA. 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci*. 10:24–30.
- Schulze K, Zysset S, Mueller K, Friederici AD, Koelsch S. 2011. Neuroarchitecture of verbal and tonal working memory in non-musicians and musicians. *Hum Brain Mapp*. 32:771–783.
- Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci*. 11:1004–1006.

- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 10:273-289.
- von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res*. 17:48-55.
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. 2005. Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci*. 17:367-376.
- Wiethoff S, Wildgruber D, Kreifelts B, Becker H, Herbert C, Grodd W, Ethofer T. 2008. Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *Neuroimage*. 39:885-893.
- Winston JS, O'Doherty J, Kilner JM, Perrett DI, Dolan RJ. 2007. Brain systems for assessing facial attractiveness. *Neuropsychologia*. 45:195-206.
- Zuckerman M, Driver RE. 1989. What sounds beautiful is good—the vocal attractiveness stereotype. *J Nonverbal Behav*. 13:67-82.
- Zuckerman M, Hodgins H, Miyake K. 1990. The vocal attractiveness stereotype—replication and elaboration. *J Nonverbal Behav*. 14:97-112.