The
British
Psychological
Society

www.bpsjournals.co.uk

Voice recognition and the posterior cingulate: An fMRI study of prosopagnosia

Stephen R. Arnott[1]*, Charles A. Heywood[2], Robert W. Kentridge[2] and Melvyn A. Goodale[1]

[1]Department of Psychology, The University of Western Ontario, London, Canada
[2]Department of Psychology, Durham University, Durham, UK

Voices, in addition to faces, enable person identification. Voice recognition has been shown to evoke a distributed network of brain regions that includes, in addition to the superior temporal sulcus (STS), the anterior temporal pole, fusiform face area (FFA), and posterior cingulate gyrus (pCG). Here we report an individual (MS) with acquired prosopagnosia who, despite bilateral damage to much of this network, demonstrates the ability to distinguish voices of several well-known acquaintances from voices of people that he has never heard before. Functional magnetic resonance imaging (fMRI) revealed that, relative to speech-modulated noise, voices rated as familiar and unfamiliar by MS elicited enhanced haemodynamic activity in the left angular gyrus, left posterior STS, and posterior midline brain regions, including the retrosplenial cortex and the dorsal pCG. More interestingly, relative to noise and unfamiliar voices, the familiar voices elicited greater haemodynamic activity in the left angular gyrus and medial parietal regions including the dorsal pCG and precuneus. The findings are consistent with theories implicating the pCG in recognizing people who are personally familiar, and furthermore suggest that the pCG region of the voice identification network is able to make functional contributions to voice recognition even though other areas of the network, namely the anterior temporal poles, FFA, and the right parietal lobe, may be compromised.

Voices, like faces, facilitate person identification (Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990). In contrast to the large amount of research devoted to the mechanisms of face recognition (c.f., Kanwisher & Yovel, 2006), there has been less work on voice recognition. Recently, however, functional imaging is revealing a network of brain regions that appear to subserve the audio component of person identification (i.e. voice identification). Although the core of this network lies along the superior temporal sulcus, results from several studies implicate additional areas in the anterior temporal

pole, fusiform gyrus and posterior cingulate gyrus (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Shah *et al.*, 2001; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005).

One of the earliest imaging studies to examine voice processing was a positron emission tomography (PET) study in which listeners attended to recorded utterances and were required to either identify the persons making each utterance or the emotion that each speaker was attempting to convey (Imaizumi *et al.*, 1997). Compared with emotion identification, it was found that speaker identification elicited greater haemodynamic activity in the temporal poles bilaterally. Subsequent studies have since confirmed this with some suggestion that the anterior portions of the right temporal pole in particular seem to play a more prominent role in the identification of voices (Belin & Zatorre, 2003; Gainotti, Barbier, & Marra, 2003; Lattner, Meyer, & Friederici, 2005; Nakamura *et al.*, 2001; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003).

Reinforcing all of these results, lesions to the anterior portions of the temporal lobe often result in an increased difficulty, if not an outright inability, to recognize people from their voice alone, even though other auditory abilities (e.g. recognizing non-human sounds such as dog barks or jingling keys) may still be preserved (Gainotti *et al.*, 2003; Gentileschi, Sperber, & Spinnler, 1999, 2001). Importantly, the anterior temporal activation associated with voice recognition is not observed when listeners attend to the verbal content of spoken sentences, suggesting that the anterior temporal area is not simply related to linguistic processing (Belin & Zatorre, 2003; Belin, Zatorre, & Ahad, 2002; von Kriegstein *et al.*, 2003). In keeping with this, a recent study by Lattner and colleagues (Lattner *et al.*, 2005) linked this area to the processing of prototypical characteristics of voices. Specifically, they found higher BOLD activity associated with natural voices as compared with unnatural (i.e. artificially lowered or raised in pitch) voices, but only in the right anterior temporal region. All of these findings are consistent with the hypothesized role that anterior portions of the temporal lobe represent a ventral auditory network that is especially important for identifying complex sounds in the environment (Alain, Arnott, Hevenor, Graham & Grady, 2001; Arnott, Binns, Grady & Alain, 2004; Arnott, Grady, Hevenor, Graham & Alain, 2005; Rauschecker & Tian, 2000).

In addition to the anterior temporal pole, other areas have been implicated in voice recognition. Thus, a recent fMRI study (von Kriegstein *et al.*, 2005) demonstrated that when listeners attended to the identity of voices (in contrast to the content of the speech), they showed activity not only in the temporal pole but also in a region of the fusiform gyrus overlapping and rostral to the fusiform face area (FFA), a region of human extrastriate cortex well documented to be face selective (Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher, McDermott, & Chun, 1997). Functional connectivity analyses of this rostral FFA area indicated that it had direct connections with the superior temporal sulcus (von Kriegstein *et al.*, 2005). Such fusiform activity cannot be entirely explained by visual imagery of the speaker's face, since similar activity associated with recognized voices has been demonstrated in the FFA of a developmental prosopagnosic patient who has never been able to process (and therefore visually imagine) faces (von Kriegstein, Kleinschmidt, & Giraud, 2006). Thus, while an intact FFA may aid in voice recognition, its absence does not necessarily ensure voice recognition deficits; a result that is in keeping with the anecdotal observation that prosopagnosic patients who have suffered damage to this area are often able to compensate for their deficit by attending to the speaker's voice.

A third and somewhat less recognized brain region often implicated in voice recognition studies is the posterior cingulate gyrus (pCG), extending into adjacent regions of the

retrosplenial cortex and the precuneus. For example, pCG activation has been reported in many of the aforementioned voice recognition studies (von Kriegstein *et al.*, 2003, 2005), in addition to several others (Nakamura *et al.*, 2001; Shah *et al.*, 2001; Stevens, 2004). A study by Shah *et al.* is particularly noteworthy since the pCG activation observed during familiar voice recognition was also found during familiar face recognition. Importantly, the familiar stimuli used in that study were obtained from people who were personally familiar to the participants, rather than those who were media famous (e.g. celebrities). These researchers proposed that the pCG receives converging input from modality-specific areas in order to assess familiarity, and that disruption to this circuit may underlie certain neurological and psychiatric disorders including prosopagnosia, phonagnosia, and Capgras delusions. In addition to its putative links with person recognition, pCG activity has also been observed in a broad range of episodic memory tasks (Wagner, Shannon, Kahn, & Buckner, 2005), as well as other recognition tests of personally familiar objects and places (Sugiura, Shah, Zilles, & Fink, 2005). Recently, the pCG has been shown to be positively correlated with memory confidence in word list recall (Moritz, Glascher, Sommer, Buchel, & Braus, 2006), consistent with an earlier autobiographical memory study showing pCG and precuneus activities in response to viewing the names of immediate friends and family members as compared with names of strangers (Maddock, Garrett, & Buonocore, 2001). Finally, there is also evidence that the area may help mediate interactions between emotional and memory-related processing (Maddock, Garrett, & Buonocore, 2003).

The present study investigates the voice recognition ability of a patient, MS, who suffered extensive damage to his voice recognition network, including bilateral temporal poles, right temporal lobe, FFA, and a large portion of his right parietal lobe as a result of a viral encephalitis developed 35 years earlier. His injuries have left him prosopagnosic, cerebrally achromatopsic and topographically amnesic. Because his posterior cingulate area was relatively intact, we were interested in determining whether or not MS has any preserved voice identification ability, and if he does, what the corresponding functional data would reveal about his person identification network.

Method

Case history

Patient MS (56-year-old, left-handed male) suffered idiopathic herpes encephalitis in 1970 resulting in extensive right hemispheric damage that left him prosopagnosic, cerebrally achromatopsic, and topographically amnesic. As can be seen in Figure 1 and as reported earlier (Heywood, Cowey, & Newcombe, 1991; Newcombe & Ratcliff, 1975; Ratcliff & Newcombe, 1982), MS has ventromedial damage bilaterally involving the lingual and the fusiform gyri. Importantly, the right temporal pole as well as the second, third, and fourth temporal gyri have suffered extensive damage. His left temporal lobe is relatively intact apart from damage to the pole, fourth temporal and parahippocampal gyri, and the mesial occipitotemporal junction. The occipital damage has left MS with a left homonymous hemianopia with macular sparing.

Since 1972, MS has been employed in the services of Remploy, a company specializing in the employment of people with disabilities. Although he is severely agnosic and unable to recognize people by their faces, he reports being able to discern identities through other non-facial cues (e.g. voice, jewellery, hair, and body size). He does not suffer from alexia, agraphia, or aphasia, and in his spare time he enjoys swimming (assisted by a coach), listening to classical music and the radio, as well as reading a newspaper.
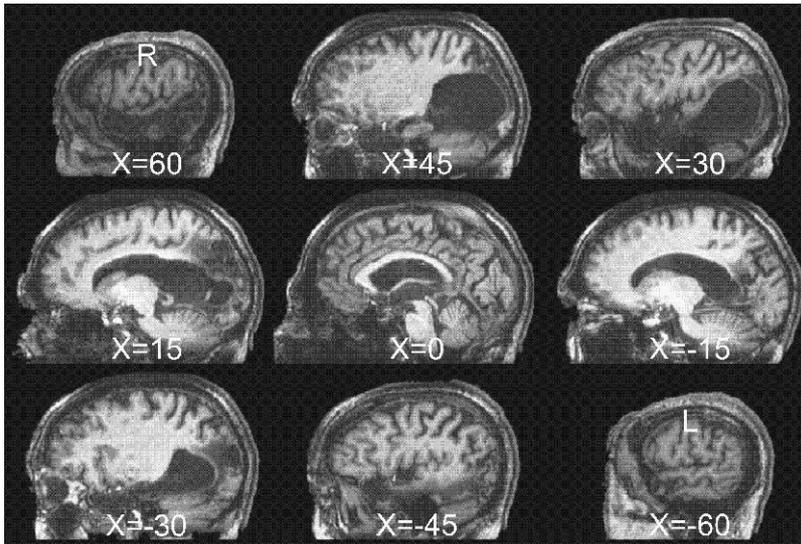
Figure 1. Anatomical T1-weighted sagittal MRI images of patient MS.

Data acquisition

Before testing, written, informed consent was obtained from the participant in accordance with ethics review committee of The University of Western Ontario. Scans were conducted using a 4-T Siemens–Varian whole-body MR scanner (Erlangen Germany; Palo Alto, CA) with a standard cylindrical quadrature head coil. Functional imaging was performed to measure the blood oxygenation level-dependent (BOLD) effect with optimal signal contrast. Seventeen, 6 mm thick axial slices were obtained. Functional scans were acquired using a navigator echo-corrected, slice-interleaved double-shot, T2*-weighted, echo-planar imaging pulse sequence (TR = 2,000 ms, TE = 15.0 ms, flip angle 40°, field of view 22.0 cm, 64 × 64 effective acquisition matrix). Functional scans were aligned to a high-resolution (1 × 1 × 1 mm) volumetric anatomical MRI that was acquired following the functional runs. Following the functional data acquisition, T1-weighted anatomical images were collected with the same slice orientation (3D magnetization-prepared turbo FLASH acquisition with inversion time (T1) = 600 ms, TR = 10 ms, TE = 5.5 ms, 22.0 × 22.0 cm field of view, 256 × 256 acquisition matrix, 3 × 3 × 6 mm voxel size, 112 axial slices, 1.5 mm thick).

Acoustic stimuli

Voice samples were obtained from five persons well known to MS, as well as from five persons who were entirely unknown to him and whose voices were expected to be unfamiliar to him. The familiar voices included those of the three male scientists who had accompanied him from the UK to Canada and to whom he had had exposure for at least the past 4 years. Two familiar voices were also selected from a sampling list of 20 'famous' voices (actors and politicians) to whom it was thought that MS would likely have had some exposure either prior to or following his brain injury. This list had included 5-second sound clips (obtained predominately from the BBC Audiointerviews Archives website; www.bbc.co.uk/bbcfour/audiointerviews/) that were devoid of any bibliographic information that might give away the speaker's identity. This list included Former

President Bill Clinton, Bob Hope, David Beckham, Former Prime Minister Edward Heath (Prime Minister of UK from 1970 to 1974, Conservative Party leader from 1965 to 1975), President George W. Bush, Alfred Hitchcock, Humphrey Bogart, Former President John F. Kennedy, Jimmy Stewart, John Cleese, John Lennon, John Wayne, Prince Charles, Princess Diana, Prime Minister Tony Blair, Sean Connery, Sir Winston Churchill, Queen Elizabeth II, and former British Prime Minister Margaret Thatcher (1979–1990). Of these voices, only the latter two were accurately identified by MS. With the exception of the two identified voices, MS showed no covert signs of even remote voice recognition (e.g. facial expressions or postural changes), even when told of the identities. After identifying the Queen's voice, MS recounted a childhood memory (later confirmed by MS's mother and sister) of the Queen waving at him as he stood roadside with his parents in a crowd watching a royal procession. Upon hearing Margaret Thatcher's voice, MS smiled and exclaimed 'Maggie!'. The content of the voice samples obtained from the three 'familiar' scientists were taken from transcriptions of the aforementioned BBC downloads. Unfamiliar voices were all British speakers. Four were acquired from recordings available on the BBC Audiointerviews Archives website (two male writers, a female poet, and a female sculptor) and the remaining voice recordings were obtained from a native Englishman in the Department of Psychology at The Univeristy of Western Ontario. The five voices used for the unfamiliar condition were chosen so that they approximated the age and gender of the familiar speakers.

For each of the 10 voice identities, five separate voice samples were obtained. Accordingly, there were a total of 50 voice sound clips, each unique in its speech content. Using Adobe Audition software (version 1.0), randomly generated, 5-second pink noise clips were modulated using speech envelope patterns randomly extracted from 10 of the 50 sound clips. All sound clips were sampled at 32 KHz, normalized and delivered over acoustically padded (30 dB attenuation) circumaural, fMRI-compatible headphones (Resonance Technology, Northridge, CA) to MS at an intensity that he judged to be comfortably audible over the background scanner noise.

Procedure

Voice familiarity task

An event-related design was used with random presentation of three types of sound clips (voices of people well known to MS, voices that MS had never heard before, and voice-modulated noise). To keep the task as straightforward as possible for MS, a two-button forced-choice task was employed. In other words, MS was instructed to press the left button if he heard a familiar voice and a right button if he heard an unfamiliar voice or a noise. Each 20-second trial began with a 300 ms pure tone signalling that a sound clip was to occur 1,700 ms later. The 5-second sound clip was followed by 13 seconds of silence. Immediately following each sound clip, MS was asked to respond with his right index finger if he judged the audio to be that of a familiar voice (he was explicitly informed of the identities of the familiar voices) and with his right middle finger if he judged the audio to be that of an unfamiliar voice or noise. Before beginning the experiment, samples of the noise stimuli were presented to MS to inform him what 'speech-modulated noise' sounded like. Five runs were presented. Each run was 5 minutes and 20 seconds in duration (with the first 20 seconds being silence) and contained 15 trials. Trial type was randomized, but over the course of the five runs, 28 familiar (three clips were repeated twice), 30 unfamiliar (five clips were repeated twice), and 17 noise clips were presented.

Functional localizer task

Following the voice familiarity task and the anatomical scan, two functional localizer runs designed to locate brain areas specialized for processing visual face, place, and object (FPO) stimuli were administered. Each run was 6 minutes and 40 seconds long and consisted of 25 blocks of intact face (four blocks), place (four blocks), or object (four blocks) stimuli, interleaved with 13 blocks of scrambled images from each category. The scrambled images served as the baseline task for the fMRI analysis. Within each block, 16 novel pictures were presented each for 850 ms with a stimulus onset asynchrony (SOA) of 150 ms. After 8 seconds of fixation, the next block began. MS was asked to fixate on a central cross (1.1°) throughout the experiment and passively view the greyscale images (11.6° × 11.6°) that were presented to him (centred 6.3° to the right of fixation to compensate for his left homonymous hemianopia).

fMRI analysis

Data processing and analyses were performed using Brain Voyager QX software (version 1.7, Brain Innovation, Maastricht, The Netherlands). After undergoing linear trend removal, high-pass filtering (three cycles), and a correction for serial correlations (i.e. the problem of inflated t-values associated with the autocorrelation of neighbouring measurement time points) using Brain Voyager, functional scans were co-registered to the last functional run before the anatomical volume and then transformed into a common stereotaxic space (Talairach & Tournoux, 1988). The data were not spatially smoothed. No motion correction was applied to the data for two reasons. First, none of the runs contained head motion that exceeded 1 mm in translation and/or 1° in rotation (as determined by Brain Voyager's 3D motion correction algorithm). Second, there is evidence indicating that motion correction algorithms can occasionally create spurious brain activations (Freire & Mangin, 2001). Baseline haemodynamic activity was taken as the 13 seconds of silence between the sound clip and the alerting cue. The reference haemodynamic-related function (HRF) was defined by Brain Voyager's two gamma haemodynamic response function (onset of curve = 0 ms, time to response peak = 5 s, response dispersion = 1, undershoot ratio = 6, time to undershoot peak = 15 s, and undershoot dispersion = 1). General linear model analyses were performed with separate predictors for each stimulus (familiar voices correctly rated as 'familiar' (FV), unfamiliar voices correctly rated as 'unfamiliar' (UV), speech-modulated noise (N), as well as voices incorrectly rated as 'familiar', voices incorrectly rated as 'unfamiliar', and finally for the cue tone at the beginning of each trial).

Statistical analysis of the fMRI data was carried out in two ways. First, a conjunction contrast of voices with speech-modulated noise (i.e. FV − N ∩ UV − N) was executed to define brain regions that were sensitive to human voices as opposed to speech-modulated noise. For this contrast, the single-voxel threshold was set to $t_{778} > 3.30$ ($p < .001$), uncorrected. A spatial cluster extent threshold was used to correct for multiple comparisons using AlphaSim (Alpha Simulations) with 1,000 Monte Carlo simulations and taking into account the entire EPI matrix (minus area outside of the brain as well as the enlarged lateral ventricles). This procedure yielded a minimum cluster size of 81 mm^3 in the acquisition space with a mapwise false positive probability of $p < .02$. A second conjunction contrast (i.e. FV − UV ∩ FV − N) was also carried out to determine brain regions most sensitive to familiar voices. For this contrast, the single-voxel threshold was set to $t_{778} > 1.97$ ($p < .05$), uncorrected and yielded a minimum cluster size of 324 mm^3 in the acquisition space with a mapwise false positive

probability of $p < .03$. In addition, some of the clusters of activation that were revealed by these conjunction contrasts underwent further comparison. In such cases, the event-related per cent-signal change averages of FV, UV, and N were extracted from the region of interest (ROI) in each experimental run. Each average was baselined using Brain Voyageur's file-based event-related average procedure (baseline was defined as the two TR values prior to the onset of the audio clip), and the per cent signal change value at each time point was extracted. A repeated measures analysis of variance (ANOVA) using the per cent signal change values at specified TRs was then carried out using each experimental run as a repeated measure.

Comparison participant

As a point of comparison, one neurologically intact male from the laboratory (34 years old, right-handed) was run in a similar experimental design. Experimental tasks and data analysis were the same as those used for MS, with the exception that different audio stimuli were employed. The five familiar voices were all laboratory members whom the participant had known for 5 years. The five unfamiliar voices were recorded from people outside of the Department of Psychology (age and gender matched to the voices used in the familiar group), whom the participant had not previously encountered before. As with the experimental design used for MS, the voice-modulated noise stimuli were created by extracting speech envelopes from familiar and unfamiliar voices, and applying them to 5-second pink noise clips. Six runs of the voice task and one run of the localizer task were completed.

Results

Behavioural

MS performed the voice recognition task competently, responding on every trial, and did not report any difficulties in hearing the stimuli or carrying out the task. Of the 28 voices from familiar people, 25 (89.3%) were accurately identified as 'familiar', and 21 of the 30 (70.0%) voices from unfamiliar people were accurately identified as 'unfamiliar'. t Tests determined that these accuracy levels were significantly above chance performance ($t = 6.6$, $p < .001$ and $t = 2.4$, $p < .05$, for FV and UV, respectively). A paired t test suggested that MS made significantly more errors for the unfamiliar voices than the familiar voices ($t = 2.12$, $p < .05$). All 17 noises were correctly identified as unfamiliar. The control participant (who also identified all noises correctly) performed as well as MS did on familiar voices, correctly identifying 28 of 32 (87.5%) as familiar, but was more accurate at identifying unfamiliar voices (96.7%). Although reaction time (RT) was not stressed and in fact was not recorded from the control participant (the control was asked to withhold his response until the sound clip had finished playing in order to avoid motor contamination of the functional data), there was a trend for MS to be faster ($t = 1.86$, $p < .07$) when rating voices as familiar (mean RT = $6,607$ ms) as opposed to unfamiliar (mean RT = $7,390$ ms). Two of the voices incorrectly identified as unfamiliar by MS on some presentations were spoken by Queen Elizabeth II, while the remaining voice was that of one of the scientists whom MS knows well. With respect to the unfamiliar speakers, all five identities were incorrectly identified as familiar at least once, two were identified as familiar twice, and one male was identified three times as familiar. Additionally, MS tended to respond faster to voices that he correctly identified as

familiar compared with those that he incorrectly identified as familiar (mean RT difference $= 1,161$ ms; $t = -1.96$, $p < .06$), providing further evidence that MS was able to differentiate familiar from unfamiliar voices.

As a final note, although they required the same button press (i.e. unfamiliar), it was clear that MS was able to differentiate 'unfamiliar' voices from the speech-modulated noise for two reasons. First, during debriefing MS reported that he had no difficulties in discriminating noise from voices. Second, MS was significantly faster (mean RT difference $= 1,251$ ms; $t = 2.21, p < .05$) and more accurate ($t = 2.64, p < .05$) on noise trials. Incidentally, these RT differences in combination with the fact that MS's response times were on average at least 1.5 seconds longer than the actual sound clips suggest that he was not simply waiting until each sound clip had ended before he made his response, but rather that some degree of stimulus processing was occurring, throughout each trial.

fMRI

As stated previously, only correct trials were analysed, and MS correctly identified 28 voices as 'familiar' and 21 as 'unfamiliar'. All 17 speech-modulated noises were correctly identified and included in the 'noise' average. In addition to the three types of sound clips (i.e. FV, UV, and N), the haemodynamic response to the cue tone was also modelled (see Figure 2). As can be seen in Figure 2a, both auditory cortices in MS responded to tonal sounds, although the extent and amount of activation were markedly reduced when compared with those of the control (Figure 2b).

Conjunction analysis 1: FV − N ∩ UV − N

Relative to speech-modulated noise, statistical comparison of MS's BOLD response to voices (both familiar and unfamiliar) revealed significant activations (i.e. $p < .001$, corrected) in the left angular and occipitotemporal gyri, as well as in the midparietal
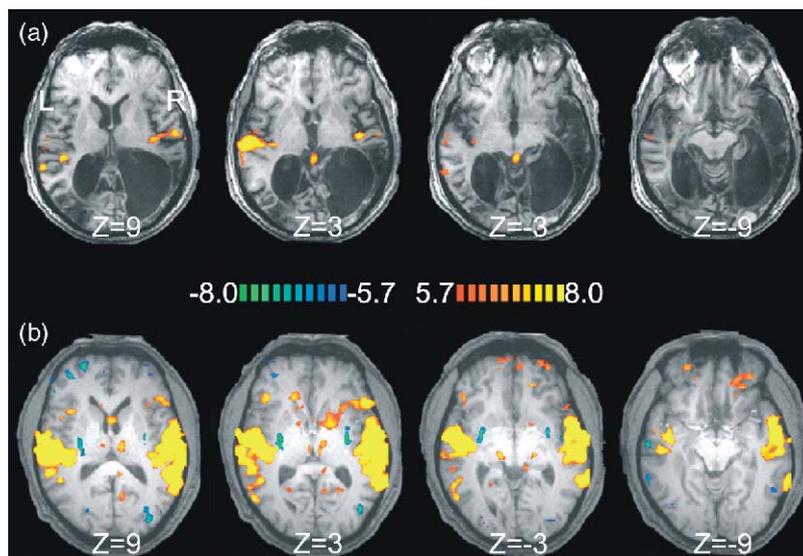
Figure 2. a) Auditory cortex activation for patient MS in response to the cue tone ($t_{778} > 5.70$, Bonferonni corrected at $p < .001$). b) Corresponding activation in the control participant ($t_{947} > 5.70$, Bonferonni corrected at $p < .001$).

regions including areas of the posterior cingulate gyrus and retrosplenial cortex (see Table 1 and Figure 3a). A cluster of activity was also found in the left posterior superior temporal sulcus (pSTS), although its $61 \, \text{mm}^3$ size fell below the $80 \, \text{mm}^3$ minimum requirement of AlphaSim at a threshold level of $p < .001$. At a less stringent threshold level of $p < .05$, however, this pSTS area increased to $595 \, \text{mm}^3$, easily attaining AlphaSim's minimum size requirement of $324 \, \text{mm}^3$. It is worth noting that at this threshold level, there was also a $167 \, \text{mm}^3$ area of voice-specific activation in the anterior pole of the left superior temporal gyrus (see dashed box in Figure 3a).

Table 1. Patient MS: familiar voices > speech-modulated noise ∩ unfamiliar voices > speech-modulated noise

Location	BA	Peak t value	x	y	z	Number of voxels; $t_{778} > 3.30$
L middle temporal (angular) gyrus	39	4.34	−60	−58	19	80*
L posterior cingulate gyrus	31	5.14	−9	−52	37	83*
L retrosplenial cortex	30/31	5.27	−3	−55	22	206*
L precuneus	7	5.35	0	−58	43	295*
L posterior superior temporal sulcus	21/37	3.94	−45	−49	−2	61
L precentral gyrus	4	5.11	−42	−4	55	122*
L superior frontal gyrus	6	3.90	−21	23	61	230*
R superior frontal gyrus	8	4.07	24	26	55	56
L medial frontal gyrus	6	4.19	−39	11	52	226*
R medial frontal gyrus	6	4.20	42	14	55	164*
R medial frontal gyrus	8	4.25	45	20	43	438*
L anterior temporal pole	*38*	*3.05*	*−48*	*21*	*−5*	*167*

*Significant at $p < .001$, corrected, when cluster size exceeds 81 as determined by AlphaSim.
Italics represent a non-significant (as determined by AlphaSim) activation at a reduced threshold of $t_{778} > 1.97$.

Examining the event-related BOLD averages in all of these areas for each of the three types of sounds suggested that the per cent signal change in the posterior cingulate at the third and fourth TR following sound onset was larger for the FV compared with UV or N conditions. An ANOVA on these per cent signal change values for each of the conditions determined the presence of a main effect of type [$F(2, 4) = 17.7, p < .001$], reflecting the fact that FV and UV signals were greater than N signal. However, despite the fact that four of the five runs showed the FV > UV pattern at the third and/or fourth TR after sound onset, a pairwise comparison of FV and UV per cent signal values failed to show a statistical difference ($p > .1$), likely reflecting, at least in part, an issue of statistical power.

By comparison, the control participant also showed significantly greater BOLD responses for voices versus noise in the posterior cingulate area and left posterior STS, as well as in several additional brain regions not seen in MS, most notably bilateral temporal poles and right superior temporal sulcus regions (see Table 2 and Figure 3b). Similar activation patterns were also found in the inferior and superior frontal regions of the control participant.

Conjunction analysis 2: FV − UV ∩ FV − N
To determine if there were any brain areas particularly activated by FV compared with UV or N, a second conjunction contrast was carried out (FV − UV ∩ FV − N). This analysis revealed significant activity in MS's medial parietal region, including the dorsal posterior cingulate and precuneus areas (see Table 3 and Figure 4a). Areas in the
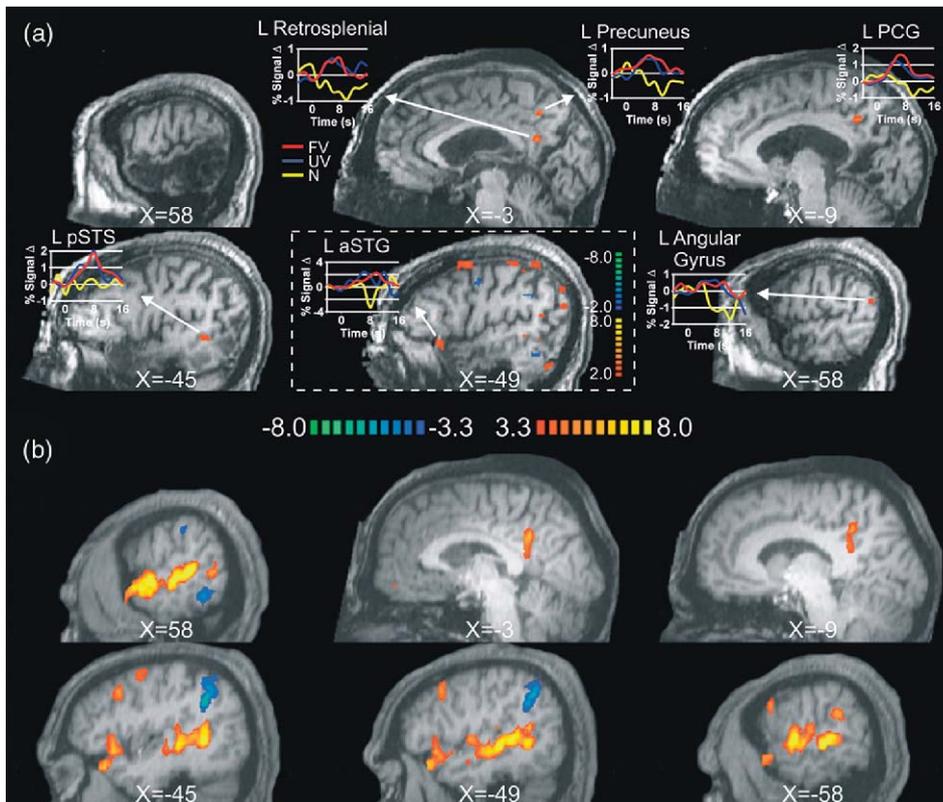
Figure 3. Voice-related activation using the conjunction contrast of familiar voices (FV) > speech-modulated noise (N) ∩ unfamiliar voices (UV) > N. a) Patient MS (t_{778} > 3.30; cluster size > 50 mm³). Event-related averages within each of the activation clusters are shown for FV (red), UV (blue), and N (yellow). Boxed image represents an area of activation at a reduced threshold t_{778} > 1.97 (cluster size = 167 mm³). b) Control participant (t_{947} > 3.30; cluster size > 300 mm³). PCG, posterior cingulate gyrus; pSTS, posterior superior temporal sulcus; aSTG, anterior superior temporal gyrus.

posterior middle temporal gyrus, superior occipital gyrus, and superior parietal lobe of the left hemisphere were also activated in MS. Similarly, the same conjunction contrast in the control participant revealed significant activity in a large region of the posterior cingulate also extending dorsally into the precuneus but, unlike for MS, inferiorly into the retrosplenial cortex as well (see Table 4 and Figure 4b). In addition, the control participant also showed statistically significant activity in the left dorsal frontal gyrus that was only observed in MS at a statistically non-significant threshold level.

FPO localizer

To assess functional activity associated with visual face processing, a passive visual experiment was administered with blocks of greyscale faces (F), places/buildings (P), and objects (O), interleaved with scrambled versions of each (baseline comparison task). Consistent with MS's reported deficit in face processing, a conjunction contrast examining face processing (i.e. F − P ∩ F − O) did not reveal any area of significant activity (see Figure 5a). The same contrast in the control participant, however, revealed a cluster of significant activity in the left fusiform gyrus (i.e. FFA; see Table 5 and

Table 2. Control: familiar voices > speech-modulated noise ∩ unfamiliar voices > speech-modulated noise

| Location	BA	Peak t value	x	y	z	Number of voxels; $t_{947} > 3.30$
L/R posterior cingulate gyrus	29/30/31	8.52	−9	−58	10	4,281*
R fusiform gyrus	37	4.77	−36	−43	−20	203*
L superior temporal gyrus and sulcus	22/21	13.59	54	−1	−2	12,345*
R superior temporal gyrus and sulcus	21/22/38/39	16.33	−55	−34	4	10,296*
R middle temporal gyrus	39	3.31	−36	−66	26	124*
R posterior temporal (including middle temporal) gyrus	21/39	6.79	−60	−49	1	549*
L lingual gyrus	18	4.40	3	−76	−23	148*
L inferior parietal lobule	40	6.80	57	−34	22	514*
L inferior frontal gyrus	47	8.23	42	33	−20	3,021*
L inferior frontal gyrus	46	5.64	42	5	49	398*
L inferior frontal gyrus	44	6.60	54	20	37	1,991*
R inferior frontal gyrus	45	7.08	−48	26	4	1,511*
L medial frontal gyrus	9	3.50	48	29	25	159*
R inferior/medial frontal gyrus	44/9	9.62	−39	14	28	2,434*
L superior frontal gyrus	9	4.69	10	53	49	136*
R superior frontal gyrus	9	6.77	−3	59	37	794*
R superior frontal gyrus	9	4.36	−18	62	28	127*
L superior frontal gyrus	8	4.07	18	50	50	189*
R dorsal frontal gyrus	8	5.56	−3	59	37	525*
R dorsal frontal gyrus	6	3.92	0	2	58	88*
L precentral gyrus	6	5.24	57	−1	40	322*
R precentral gyrus	6	4.32	−51	5	34	148*
R precentral gyrus	6	6.69	−45	−1	49	1,593*

*Significant at $p < .001$, corrected, when cluster size exceeds 81 as determined by AlphaSim.

Figure 5b) as well as the right precuneus and left supramarginal gyrus. Based on previous reports of FFA activation in response to voice identification (von Kriegstein *et al.*, 2005), voice-related BOLD activity was re-examined in the control participant within the left fusiform area. Running a repeated measures ANOVA on three conditions (FV, UV, and N) at third, fourth, fifth, and sixth TR after sound onset resulted in a main

Table 3. Patient MS: familiar voices > unfamiliar voices ∩ familiar voices > speech-modulated noise

Location	BA	Peak t value	x	y	z	Number of voxels; $t_{778} > 1.97$
L middle temporal gyrus/superior occipital gyrus	39/19	2.93	−27	−61	22	348*
L superior parietal lobule	7	3.17	−30	−58	37	311
L/R dorsal posterior cingulate gyrus	31/7	3.52	0	−49	43	1,810*
R superior frontal gyrus	6	3.75	3	5	64	695*

*Significant at $p < .05$, corrected, when cluster size exceeds 324 as determined by AlphaSim.
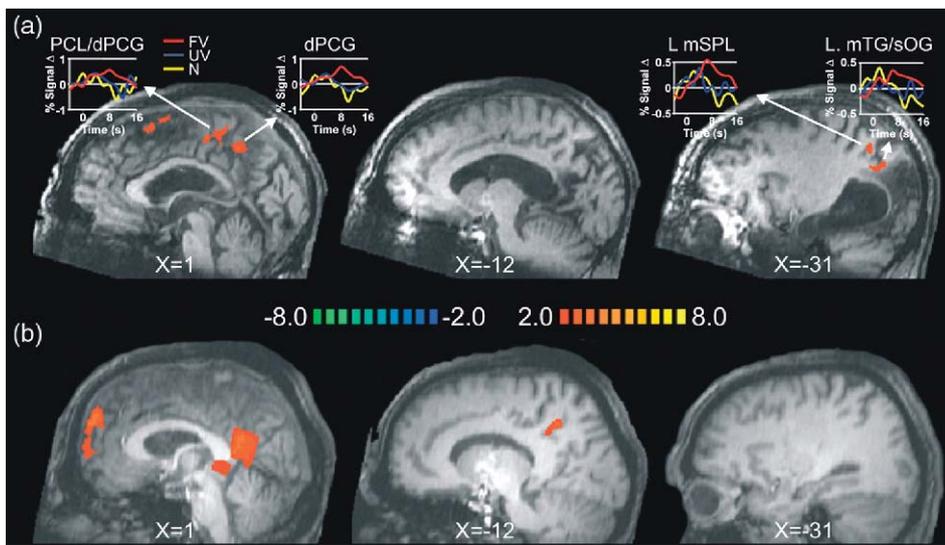
Figure 4. Familiar voice-related activation using the conjunction contrast of familiar voices (FV) > unfamiliar voices (UV) ∩ FV > speech-modulated noise (N). a) Patient MS ($t_{778} > 2.00$; cluster size > 300 mm^3). Event-related averages within each of the activation clusters are shown for FV (red), UV (blue), and N (yellow). b) Control participant ($t_{947} > 2.00$; cluster size > 324 mm^3). PCL, paracentral lobule; dPCG, dorsal posterior cingulate gyrus; mSPL, medial superior parietal lobe; mTG, middle temporal gyrus; sOG, superior occipital gyrus.

effect of condition [$F(2, 12) = 4.61$, $p < .05$]. Pairwise comparisons revealed that the per cent signal BOLD responses to both voice stimuli were significantly greater than that of the noise stimuli (both $p < .05$), but that the BOLD response to the two types of voice stimuli were not significantly different from one another ($p > .10$).

Discussion

MS, a person with extensive damage to anterior temporal poles, fusiform and right parietal regions, nevertheless demonstrates some preserved ability to recognize voices. An fMRI study revealed that when he was engaged in such a task, activation was

Table 4. Control: familiar voices > unfamiliar voices ∩ familiar voices > speech-modulated noise

Location	BA	Peak t value	x	y	z	Number of voxels; $t_{947} > 1.97$
L/R posterior cingulated including:	26/29/23	4.14	0	−43	10	5,123*
L posterior cingulated/precuneus	31/7		12	−49	35	
L posterior cingulated/precuneus	31/7		5	−39	30	
Bilateral retrosplenial cortex	30		−3	−36	0	
L dorsal frontal gyrus	9	4.55	0	56	31	2,136*
L dorsal frontal gyrus	10	3.00	0	50	4	253
R thalamus		2.65	−3	−4	4	286
R dorsal entorhinal area	34	3.02	−18	−19	−11	213

*Significant at $p < .05$, corrected, when cluster size exceeds 324 as determined by AlphaSim.
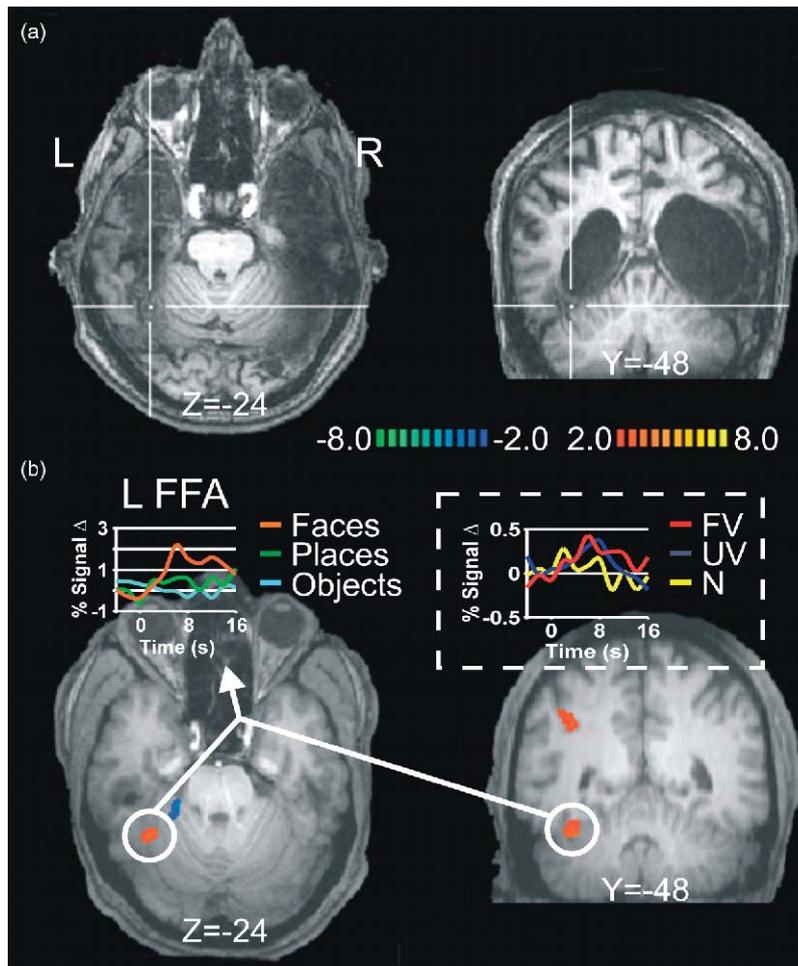
Figure 5. Cortical areas associated with visual face processing using a conjunction contrast of faces > places ∩ faces > objects. a) Patient MS (t_{778} > 2.00; cluster size >200 mm^3). b) Control participant (t_{947} > 2.00; cluster size >200 mm^3). Event-related averages within the activation cluster are shown for faces (orange), places (green), and objects (cyan). The dashed box contains the corresponding event-related averages within this same brain region during the voice recognition experiment. FV, familiar voices (red), UV, unfamiliar voices (blue), and N, noise (yellow). FFA, fusiform face area.

observed principally in the posterior temporal regions of the left hemisphere, and dorsal posterior cingulate and medial parietal regions. In conjunction with results from other voice processing experiments (Belin, Fecteau, & Bedard, 2004), computational approaches to auditory processing (Griffiths *et al.*, 2007), as well as from our comparison participant, we speculate that the spared region within the left STS, in particular, is what enabled MS to accomplish the early stages of voice processing (e.g. the ability to extract the 'voice' quality from the sounds). In addition, because MS suffers a complete lesion of the FFA as confirmed by anatomical and functional MRI, our results clearly show that an intact FFA is not essential for all types of voice recognition. What is remarkable is the fact that MS was able to perform as well as he did at discriminating the voices considering the amount of damage inflicted to his voice identification

Table 5. Control: faces > places ∩ faces > objects

Location	BA	Peak t value	x	y	z	Number of voxels with $t_{947} > 1.97$
L supramarginal gyrus/inferior parietal lobule	40	3.39	36	−49	34	619*
L fusiform gyrus	37	3.48	33	−47	−26	352*
R parahippocampal gyrus	37/19	−4.68	24	−55	−14	370*
R parahippocampal gyrus	36	−4.09	21	−31	−24	345*
R precuneus	19	3.00	−36	−76	40	583*
R inferior frontal gyrus	45/46	3.01	−36	29	10	253
R inferior frontal gyrus	44	2.92	−48	8	31	220

*Significant at $p < .05$, corrected, when cluster size exceeds 324 as determined by AlphaSim.

network, particularly in the anterior temporal regions. The latter have been consistently shown to be activated during voice identification tasks (Belin & Zatorre, 2003; Belin *et al.*, 2002; Lattner *et al.*, 2005; von Kriegstein *et al.*, 2003), and in many cases, damage to these areas greatly degrades, if not eliminates, a listener's ability to identify voices (Gainotti *et al.*, 2003; Gentileschi *et al.*, 1999, 2001). It is possible that a small area of spared cortex in MS's left anterior temporal pole was able to make a functional contribution to his voice identification ability. Indeed, greater BOLD activation was observed in that area in response to familiar voices. Even though the cluster size of this activation was well below the corrected statistical threshold, the dismissal of this activation should be tempered by the fact that there is a severe reduction in grey matter in this region.

Having said that, it is important to note that although MS performed well above chance at recognizing the familiar voices that were used in this study, one would be hesitant to claim that MS has normal voice recognition ability. With the exception of two speakers, a test of 20 famous voices evoked not even a hint of recognition from MS. In some respects, this result is not at all surprising given the extent of his brain injury and what is known about the role of the temporal lobe in voice processing (Assal, Aubert, & Buttet, 1981; Assal, Zander, Kremin, & Buttet, 1976; Belin *et al.*, 2002; Lattner *et al.*, 2005) and the right parietal lobe's involvement in famous person recognition (Van Lancker, Kreiman, & Cummings, 1989). The preserved ability that MS exhibits then appears to be related to the particular voices used in this experiment. Indeed, most of the 'familiar' voices used in the present experiment were those of people with whom MS was personally familiar. This finding certainly is in accordance with the results from Shah and colleagues (2001) as well as a group lead by Nakamura (Nakamura *et al.*, 2001). The former found that pCG areas, including the retrosplenial cortex and precuneus, were more active for both the faces and voices of friends and relatives (as compared with appropriate 'unfamiliar' comparison stimuli), while the latter found left precuneus activation specific to personally familiar voices. Some studies have shown activation in this region with the faces of famous people (Gorno Tempini *et al.*, 1998; Leveroni *et al.*, 2000), leading to the suggestion that voice- and face-related activation in this region occurs when there is some emotional connection with the individual (Maddock *et al.*, 2003). It is interesting to note that although MS could not identify the voices of most of the famous people he was presented with in the behavioural study, the two voices he did recognize, those of Margaret Thatcher and Queen Elizabeth, belonged to the two individuals with whom it could be argued he had the most emotional connection.

It could be argued that the greater BOLD activation in pCG and left posterior middle temporal gyrus/superior occipital gyrus associated with familiar voices reflected greater attention to these voices as opposed to unfamiliar ones. This explanation is unlikely, however. Although MS's accuracy did not significantly differ between voices rated as familiar and unfamiliar, his reaction times were significantly slower in response to the unfamiliar. Thus, if anything, it would seem that more attention (by virtue of more effort) was devoted to unfamiliar than familiar voices. By the same token, the reaction time difference between the noise and the voice stimuli does legitimately raise the possibility that some of the activation observed in the contrast of voices with noise (e.g. parietal and frontal areas) may have been attentional in nature. While research has shown that activation in at least the anterior portions of the STS can be modulated by attention to voices (Alho *et al.*, 2006), it seems unlikely that our pSTS activation can be completely accounted for by attention, given the wealth of evidence implicating the STS as a crucial early stage in the analysis of voices (Belin *et al.*, 2000, 2004; Griffiths *et al.*, 2007; Rämä & Courtney, 2005; von Kriegstein & Giraud, 2004; von Kriegstein *et al.*, 2005).

The results of the neurologically intact participant are also interesting. For this individual, voices were found to evoke a significant increase in the BOLD signal relative to speech-modulated noise in an area of the brain that was selective for visual face stimuli (as determined by a FFA functional localizer task). This result is therefore consistent with an earlier study reporting that familiar voice identification evoked cross-modal activity in the FFA (von Kriegstein *et al.*, 2005). It is unclear in the present study whether the activation reflects explicit visual imagery of the speaker's face or not (O'Craven & Kanwisher, 2000). Although it is a possibility that cannot be completely discounted, there are a few points that argue against this. First, when debriefed after the experiment, the participant reported only as being 'aware of a face a couple of times' throughout the experiment, and this was only when the familiar voices were presented. Second, although there was some activation in the left lingual gyrus, there was no significant voice-specific activation in the calcarine sulcus as has been reported in other face imagery tasks (Ishai, Haxby, & Ungerleider, 2002). Third, the BOLD difference in FFA was present not only for familiar voices whose face was known but also for unfamiliar voices whose faces were not known to the listener. If imagery was the only underlying factor involved in the FFA activation, one might expect to have observed differences between the amounts of activation for faces that could easily be imagined versus those that could not be easily imagined. Finally, a case study of a patient who has never been able to process faces (i.e. a congenital prosopagnosic) also showed activity in the FFA (von Kriegstein *et al.*, 2006), which makes the point that voice-related processing in the FFA can be dissociated from explicit imagery. However, whatever is going on in the FFA when people recognize voices, the fact that MS continues to show some voice recognition without an FFA suggests that this region is not absolutely essential for voice recognition.

Although we did not ask MS explicitly to identify the voices during the experiment, we are confident that he can identify them. During debriefing, several sounds clips were played to him again and he was able to name all the familiar speakers accurately. One of these voices in particular elicited a virtually instantaneous recognition that was underscored by smiles and postural changes (e.g. sitting bolt upright). One interesting question, of course, is whether MS's voice recognition ability reflects some partial sparing of normal voice recognition or some atypical ability that developed in the years following the brain insult. The fact that he was able to identify his parents by voice

immediately after the precipitating event (confirmed by MS's mother and sister) suggests that the former possibility is at least plausible.

Acknowledgements

References

Alain, C., Arnott, S. R., Hevenor, S. J., Graham, S., & Grady, C. L. (2001). 'What' and 'where' in the human auditory system. *Proceedings of the National Academy of Science of the United States of America*, *98*, 12301–12306.

Alho, K., Vorobyev, V. A., Medvedev, S. V., Pakhomov, S. V., Starchenko, M. G., Tervaniemi, M., *et al.* (2006). Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. *Brain Research*, *1075*(1), 142–150.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage*, *22*, 401–408.

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience*, *17*, 1–13.

Assal, G., Aubert, C., & Buttet, J. (1981). Cerebral asymmetry and voice recognition. *Revista de Neurologia*, *137*(4), 255–268.

Assal, G., Zander, E., Kremin, H., & Buttet, J. (1976). Voice discrimination in patients with cerebral cortical lesions. *Schweizer Archiv für Neurologie, Neurochirurgie und Psychiatrie*, *119*(2), 307–315.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, *14*(16), 2105–2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(Pt 3), 305–327.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(Pt 3), 361–380.

Freire, L., & Mangin, J. F. (2001). Motion correction algorithms may create spurious brain activations in the absence of subject motion. *Neuroimage*, *14*(3), 709–722.

Gainotti, G., Barbier, A., & Marra, C. (2003). Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain*, *126*(Pt 4), 792–803.

Gentileschi, V., Sperber, S., & Spinnler, H. (1999). Progressive defective recognition of familiar people. *Neurocase*, *5*(5), 407–424.

Gentileschi, V., Sperber, S., & Spinnler, H. (2001). Crossmodal agnosia for familiar people as a consequence of right infero-polar temporal atrophy. *Cognitive Neuropsychology*, *18*(5), 439–463.

Gorno Tempini, M., Price, C., Josephs, O., Vandenberghe, R., Cappa, S., Kapur, N., *et al.* (1998). The neural systems sustaining face and proper-name processing. *Brain*, *121*(11), 2103–2118.

Griffiths, T. D., Kumar, S., Warren, J. D., Stewart, L., Stephan, K. E., & Friston, K. J. (2007). Approaches to the cortical analysis of auditory objects. *Hearing Research*, *229*(1–2), 46–53.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562.

Heywood, C. A., Cowey, A., & Newcombe, F. (1991). Chromatic discrimination in a cortically colour blind observer. *European Journal of Neuroscience*, *3*(8), 802–812.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., *et al.* (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport*, *8*(12), 2809–2812.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: Effects of memory and attention revealed by fMRI. *Neuroimage*, *17*(4), 1729–1741.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *361*(1476), 2109–2128.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, *24*(1), 11–20.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *Journal of Neuroscience*, *20*(2), 878–886.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*(3), 667–676.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, *18*(1), 30–41.

Moritz, S., Glascher, J., Sommer, T., Buchel, C., & Braus, D. F. (2006). Neural correlates of memory confidence. *Neuroimage*, *33*, 1188–1193.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., *et al.* (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia*, *39*(10), 1047–1054.

Newcombe, F., & Ratcliff, G. (1975). *Agnosia: A disorder of object recognition*. Paper presented at the Les Syndromes de Disconnexion calleuse chez l'Homme., Hopital neurologique de Lyon, Lyon.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Rämä, P., & Courtney, S. M. (2005). Functional topography of working memory for face or voice identity. *Neuroimage*, *24*(1), 224–234.

Ratcliff, G., & Newcombe, F. (1982). Object recognition: Some deductions from the clinical evidence. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 147–171). London: Academic Press.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11800–11806.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., *et al.* (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain*, *124*, 804–815.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research*, *18*(2), 162–171.

Sugiura, M., Shah, N. J., Zilles, K., & Fink, G. R. (2005). Cortical representations of personally familiar objects and places: Functional organization of the human posterior cingulate cortex. *Journal of Cognitive Neuroscience*, *17*(2), 183–198.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, *11*(5), 665–674.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, *17*(1), 48–55.

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*(2), 948–955.

von Kriegstein, K., Kleinschmidt, A., & Giraud, A. L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex*, *16*(9), 1314–1322.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, *17*(3), 367–376.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445–453.