An fMRI Study of Verbal Self-monitoring: Neural Correlates of Auditory Verbal Feedback

The ability to recognize one's own inner speech is essential for a sense of self. The verbal self-monitoring model proposes that this process entails a communication from neural regions involved in speech production to areas of speech perception. According to the model, if the expected verbal feedback matches the perceived feedback, then there would be no change in activation in the lateral temporal cortices. We investigated the neural correlates of verbal self-monitoring in a functional magnetic resonance (fMRI) study. Thirteen healthy male volunteers read aloud presented adjectives and heard their auditory feedback which was experimentally modified. Decisions about the source of the feedback were made with a button-press response. We used a 'clustered' fMRI acquisition sequence, consisting of periods of relative silence in which subjects could speak aloud and hear the feedback in the absence of scanner noise, and an event-related design which allowed separate analysis of trials associated with correct attributions and misattributions. Subjects made more misattribution responses when the feedback was a distorted version of their voice. This condition showed increased superior temporal activation relative to the conditions of hearing their own voice undistorted and hearing another person's voice. Furthermore, correct attributions during this condition were associated with greater temporal activation than misattributions. These findings support the self-monitoring model as mismatches between expected and actual auditory feedback were associated with greater temporal activation.

Keywords: corollary discharge, schizophrenia, self-monitoring, speech, temporal cortex

Introduction

The ability to distinguish one's own speech from other peoples' is essential for normal human interactions. When one speaks aloud, one is aware of the intention to speak and the act of speaking, and one receives the feedback of hearing one's own voice (Levett 1983, 2001; Frith, 1992). It is widely held that the control of somatic movement involves internal models which predict and correct the motor command prior to its effective output as a physical action (Wolpert et al., 1995). The neural components of the somatic monitoring model involve some form of a 'corollary discharge' (Helmholtz, 1866; Sperry, 1950; McCloskey et al., 1974; McCloskey and Torda, 1975), 'efference copy' (von Holst, 1954) or 'forward model' (Wolpert et al., 1995) which accompanies motor intent, as well as sensory feedback loops (Evarts, 1971; Wolpert et al., 1995). Similarly, some form of a corollary discharge from areas involved in verbal generation may inform areas in involved in speech perception that impending speech has arisen from an internal source (Ploog, 1979). The notion that the generation of speech and inner thoughts is analogous to the production of motor movements involving internal monitors and feedback loops was proposed by Freud and Hughlings Jackson over a century ago (Fuster, 1997). Such a verbal self-monitoring system may contribute to the conscious awareness of thought and play a role in distinguishing self from non-self (Feinberg, 1978; Frith, 1992).

In the verbal self-monitoring model, Frith (1992) proposed that the anterior cingulate cortex has a key role in controlling vocalization through Broca's area while also producing a corollary discharge to Wernicke's area, thereby modifying speech perception, and is itself modified by the striatal loop which includes the thalamus. According to the model, if the speech signal predicted on the basis of the motor output matches what is actually perceived, then there is no change in activation in the lateral temporal cortices (Frith et al., 1998). Electrophysiological data from monkeys (Ploog, 1979) and humans (Muller-Preuss and Ploog, 1981; Creutzfeldt et al., 1989; Ford et al., 2001) support this contention as feed forward information about self-generated verbal output produces a net suppressive effect on temporal cortical activation. With overt speech, microelectrode recordings in the middle and superior temporal gyri have revealed suppression of activity in up to one-third of the neuronal population (Creutzfeldt et al., 1989), which may even precede (Creutzfeldt et al., 1989) or occur within milliseconds of speaking (Curio et al., 2000).

An experimental paradigm which engages verbal self-monitoring entails subjects speaking aloud and hearing their auditory verbal feedback through headphones in real-time, in which the verbal feedback is experimentally modified such that it may be their voice, an altered version of their voice, or substituted by another individual voice (McGuire et al., 1996; Johns et al., 2001). In an [15O]H2O positron emission tomography (PET) study, McGuire et al. (1996) found that when subjects spoke aloud but heard a pitch distorted version of the subject's own voice or another person's voice, there was greater activation in the lateral temporal cortices. A methodological drawback of the PET study was that it employed a blocked design. When individuals speak in the presence of distorted feedback they may correctly recognize it as self-generated, be uncertain of its origin or misidentify it as 'alien' (Johns et al., 2001). The activation observed during a block may thus reflect an average across trials associated with different attributions of source. As subjects were not required to identify the source of the feedback (McGuire et al., 1996), the extent to which this affected the findings is unclear. A further disadvantage of presenting a series of the same types of trials in a block is that this may have reduced the demands on self-monitoring as after a few trials subjects may have concluded that all those in a given block were of the same type.
The present study sought to investigate the neural correlates of verbal self-monitoring using functional magnetic resonance imaging (fMRI) as it has greater spatial and temporal resolution than PET. We used an event-related design with trials from different feedback conditions presented randomly so that subjects could not anticipate successive events. The subject’s perception of the source of the feedback was measured online via a button press, which permitted the categorization of activations associated with each trial according to response accuracy. However, a key disadvantage of fMRI is the production of high amplitude acoustic noise (~110 dB) during image acquisition (Amaro et al., 2002), which makes it difficult for subjects to hear auditory stimuli. Thus, a ‘clustered’ sequence was used in which a ‘silent’ period is interleaved with the acquisition of brain images (Eden et al., 1999; Hall et al., 1999), and single trials of the task can be performed while the scanner is transiently silent. It was during these ‘windows’ that trials of the self-monitoring task was performed, which allowed subjects to speak and hear the overt verbal feedback in the absence of scanner noise.

We expected that the experimental conditions in which subjects spoke aloud but heard their own voice which was distorted by pitch (self-distorted condition) or replaced by another person’s voice (alien-undistorted) would be associated with a mismatch between what was expected as auditory feedback (the subject’s own voice undistorted) and the actual feedback. These conditions would thus be associated with greater lateral temporal activation relative to the condition in which the feedback was of their own voice undistorted (self-undistorted condition). As subjects have more difficulty in recognizing their own voice when it has been distorted than in recognizing another person’s voice (Johns et al., 2001), the self-distorted condition appears to place more demands on verbal self-monitoring than the alien-undistorted condition. We thus expected that the former condition would be associated with greater lateral temporal activation than the latter.

With the design of the present study, it was possible to analyse separately the attribution responses for each condition. We were particularly interested in the self-distorted condition as this feedback shows the greatest difference in external misattributions between individuals with schizophrenia and healthy comparison subjects (Johns et al., 2001). We sought to examine the neural correlates associated with correct and incorrect attribution responses. Our hypothesis for the neural correlates associated with this decision-making process is more tentative. During this condition, when individuals hear a distorted version of their own voice, there is a mismatch between what is expected and what is actually perceived, and it is hypothesized that the mismatch is associated with increased lateral temporal activation (McGuire et al., 1996). When subjects correctly recognize the source of the feedback, there is less of a mismatch between the expected and actual feedback, and thus less of a difference in activation in the lateral temporal cortices. Alternatively, as this decision is made after subjects have spoken and heard the auditory feedback in the present study, it is a cognitive process which occurs following the implicit self-monitoring process and may additionally engage prefrontal cortical regions (Frith, 1992).

Materials and Methods

Subjects
All subjects were dextral (Annett, 1970), native English-speaking males, free of current and past psychiatric and medical disorders. Thirteen healthy male individuals were recruited, age 32.3 ± 8.4 years (mean ± SD), mean IQ 106.2 ± 15.0 (Quick Test; Ammons and Ammons, 1962). The study was approved by the Institute of Psychiatry and South London and Maudsley NHS Trust Ethical Committee. All subjects provided written informed consent.

fMRI Data Acquisition
T2*-weighted volume images were acquired on a 1.5 T Neuro-optimized GE scanner (General Electric, Milwaukee, WI) at the Maudsley Hospital (London, UK). Twelve non-contiguous axial planes (7 mm thickness, slice skip 1 mm) parallel to the anterior commissure-posterior commissure line were collected over 1.1 s using a ‘clustered’ acquisition (12–14) (TE = 40 ms, 70° flip angle), which created a relative silent period of 2.15 s for each stimulus within a TEs of 3.25 s of the inter-stimulus interval of 16.25 s (Fig. 1). The clustered acquisition sequence was used in order to minimize any effects of susceptibility artifacts.
Figure 1. The verbal self-monitoring task was an event-related fMRI design. Brain images were acquired in a ‘clustered’ acquisition over 1.1 s, with 2.15 s of relative silence during which the word stimulus was presented, subjects read the word aloud and heard the verbal feedback through headphones. Five brain volumes were acquired for each trial ($T_R = 3.25 \text{s}$), and a stimulus was presented every 16.25 s. The BOLD response for each trial is represented (continuous line) as well as the contribution of scanner noise (dotted line) which is constant following the first trial.

In order to sample the distribution of SSQ ratio under the null hypothesis that observed values of SSQ ratio were not determined by the experimental design (with minimal assumptions), the time series at each voxel was permuted using a wavelet-based resampling method (Bullmore et al., 2001). This process was repeated 10 times at each voxel, and the data combined over all intracerebral voxels, resulting in 10 permuted parametric maps of the SSQ ratio at each plane for each subject. Combining these data yielded the distribution of SSQ ratio under the null hypothesis.

Group Mapping
The observed and randomized SSQ ratio maps were transformed into standard space by a two-stage process involving a rigid body transformation of the fMRI data onto a high-resolution inversion recovery image from the same subject, followed by an affine transformation onto a Talairach and Tournoux (1988) template. A brain activation map was produced for each condition by testing the median observed SSQ ratio (median values were used to minimize outlier effects) at each intracerebral voxel in standard space (Brammer et al., 1997) against a critical value of the permutation distribution for median SSQ ratio ascertained from the spatially transformed wavelet-permuted data. Only activations in-phase with the BOLD response were included in the analysis. In order to increase sensitivity and reduce the number of statistical comparisons, hypothesis testing was carried out at the cluster level (Bullmore et al., 1999b). The probability of occurrence of clusters under the null hypothesis was determined using the distribution of median SSQ ratios computed from spatially transformed data obtained from wavelet permutation of the time series at each voxel. Image-wise expectation of the number of false-positive clusters under the null hypothesis was set such that <1 false-positive activated clusters were expected, at a typical $P$ value of <0.01, and cluster volumes were a minimum of 5 voxels.

Between Condition Differences
Analysis of variance was carried out on the SSQ ratio maps in standard space by computing the difference in mean SSQ ratio between conditions and using a null distribution obtained by random permutation of condition membership and re-computation of the mean SSQ ratio associated with overt speech during fMRI scanning. With the clustered acquisition sequence, there was no fMRI scanning during the period in which subjects were required to make an overt speech response. The acquisition sequence began when the subject had returned to the usual resting, mouth closed state. We have found that the ‘clustered’ acquisition sequence can deal with the problem of speech movements in a better manner than other acquisition techniques (Amaro et al., 2002). Initiation of the acquisition ‘triggered’ the stimulus presentation software program to show the first stimulus for each set of stimuli. A total of 540 volumes were acquired for each subject, which were acquired as three sets of 180 volumes each.

Behavioural Response Analysis
A repeated-measures analysis of variance with within factors of source of feedback (self, alien) and level of distortion (none, distortion) was performed for each of the possible responses: correct attributions, misattribution errors and unsure responses.

fMRI Data Analysis
Individual Subject Mapping
The initial six volumes (following the word ‘begin’) and the final four volumes of each set were removed as the first and last trials were not fully sampled. The remaining data from the three sets for each scan session were concatenated. The data were realigned to the initial volume (Bullmore et al., 1999a) to minimize motion related artefacts, and smoothed by a 2D Gaussian filter (full-width half-maximum = 7.2 mm). Responses to the experimental conditions were detected by time-series analysis using Poisson functions (peak responses at 4 and 8 s) to model the blood oxygenation level dependent (BOLD) response. The analysis involved each condition being convolved separately with the Poisson functions to yield two models of the expected haemodynamic response. The weighted sum of these two convolutions that gave the best fit (determined by least-squares analysis) to the time series at each voxel was computed. This was the ratio of the sum of squares of deviations from the mean intensity value due to the model (fitted time series) divided by the sum of squares due to the residuals (original time series minus model time series), called the sum of squares quotient (SSQ) ratio.
difference. Again, maps were computed using cluster-level statistics with an expected Type I error of <1 false-positive cluster for each map.

Comparisons were first performed to examine the effect of the feedback conditions on the neural substrates of self-monitoring. The four conditions were presented in a factorial design with two levels of source and two levels of distortion so that subjects would not be able to make a systematic decision about the source of the feedback. However, according to the self-monitoring model, only one condition matches the expected feedback (self-undistorted), while the remaining three conditions (self-distorted, alien-undistorted, and alien-distorted) represent mismatches between expected and actual feedback. Thus, the self-undistorted feedback condition was contrasted with (i) self-distorted and (ii) alien-undistorted feedback conditions. These contrasts also replicated the comparisons made in the PET study (McGuire et al., 1996). The next set of contrasts examined the second hypothesis with direct comparisons of the self-distorted feedback condition with the alien-undistorted condition.

The event-related design of the present experiment allowed additional evaluation of individual trials, which had not been possible in the PET study (McGuire et al., 1996). These contrasts examined the activations associated with correct attributions and incorrect attributes (misattributions) of the feedback source. To examine the neural correlates associated with the accuracy of attributions, comparisons were made between correct responses and (incorrect) misattributions during self-distorted feedback. In order to assess whether the differences may be related to the attribution itself, additional contrasts were performed between correct self-attributions during self-undistorted and self-distorted feedback conditions and between correct self-attributions during self-undistorted feedback and correct other-attributions during the alien-undistorted condition.

**Results**

**Behavioural Data**

Subjects had little difficulty in correctly identifying their own voice during the self-undistorted feedback condition (correct responses = 95.8 ± 8.3%, misattributions = 0.9 ± 2.4%, unsure = 1.3 ± 3.7%), but during the self-distorted feedback condition they made substantial external misattributions and unsure responses (correct responses = 50.4 ± 34.4%, misattributions = 26.8 ± 27.5%, unsure = 22.3 ± 21.0%). During the alien-undistorted feedback condition, subjects correctly attributed the source to someone else on most trials, but also made self-misattributions errors (when they misidentified their own voice) and a few unsure responses (correct responses = 63.4 ± 23.3%, misattributions = 22.7 ± 19.8%, unsure = 10.5 ± 16.3%). A similar pattern of responses was evident during the alien-undistorted feedback condition (correct responses = 55.8 ± 27.5%, misattributions = 22.7 ± 18.5%, unsure = 18.0 ± 16.7%) (Fig. 2).

For correct responses, there was a main effect of distortion ($F = 49.6, df = 1.12, P < 0.001$) with more errors made during the distorted trials, and a significant distortion by source interaction ($F = 5.9, df = 1.12, P < 0.03$) with fewer correct responses made during the self-distorted relative to the self-undistorted feedback condition but no difference with distortion during the alien feedback conditions. For misattributions, there was a main effect of distortion ($F = 17.3, df = 1.12, P < 0.001$) but not of source ($F = 3.1, df = 1.12, P = 0.11$), and a trend towards a distortion by source interaction ($F = 4.2, df = 1.12, P = 0.06$). Again, this reflected more external misattributions during the self-distorted feedback condition relative to self-undistorted feedback, but no difference in the alien conditions. For unsure responses, there was a main effect of distortion ($F = 17.4, df = 1.12, P < 0.001$) but not of source ($F = 1.4, df = 1.12, P = 0.26$), and a trend towards a distortion by source interaction ($F = 3.8, df = 1.12, P = 0.07$) with more unsure responses during the self-distorted as compared with the self-undistorted feedback condition.

**fMRI Data**

**Effects of Speaking Aloud during Individual Feedback Conditions**

In order to examine the regions associated with reading aloud with auditory verbal feedback, contrasts were made of the individual conditions with the baseline condition of viewing a blank screen and not speaking. All the individual conditions were associated with activations that included the anterior and posterior cingulate gyri, ventrolateral and inferior frontal cortices, bilateral lateral temporal cortices, parietal and occipital cortices, striatum, thalamus and cerebellum. Full list of coordinates and figures are available as supplementary material.

**Contrasts between Conditions**

**Self-undistorted versus Self-distorted Feedback.** The self-undistorted condition showed greater activity relative to the self-undistorted condition in the lateral temporal cortices bilaterally: in the left temporal cortex with one cluster extending from the middle temporal gyrus [Broadmann’s area (BA) 21, Talairach and Tournoux coordinates {$x, y, z$} = {-53, -26, -7}]] to the superior temporal gyrus (BA 42 {-57, -17, 20}], cluster size 142 voxels); a more anterior cluster in the superior temporal gyrus (BA 22 {-57, -4, 4} to BA 42 {-53, -4, 9}, 61 voxels), the posterior middle temporal gyrus (BA 37 {-50, -56, 4}, 11 voxels) and the right temporal cortex (BA 22 {47, 10, -2}, 8 voxels) and another cluster in the right posterior inferior temporal gyrus (BA 37 {36, -67, 4}, 6 voxels). Additional regions which showed greater activity with self-distorted feedback were the anterior cingulate (BA 24), posterior cingulate (BA 31) and right inferior frontal (BA 47) gyri, primary occipital cortex (BA 18), putamen and brainstem.
Conversely, the self-undistorted condition was associated with greater engagement of the left temporal pole (BA 38, from {-36, 10, -18} to {-36, 10, -13}, 27 voxels), left middle temporal gyrus (BA 21 {-53, -13, -7}, 37 voxels), right superior temporal gyrus (BA 42 {53, -13, 9}, 118 voxels) and left thalamus relative to self-distorted feedback (Fig. 3). Full coordinates are available as supplementary material.

**Self-undistorted versus Alien-undistorted Feedback.** Alien-undistorted feedback showed greater activation relative to self-undistorted feedback in the lateral temporal cortices bilaterally: two clusters in the left temporal cortex, middle temporal gyrus (BA 21 {-57, -1, -2}, 22 voxels, and from {-47, -46, 4} to {-47, -46, 9}, 49 voxels) and superior temporal gyrus (BA 22/42 {-57, -17, 4} to {-57, -20, 26}, 188 voxels), and in the right middle temporal gyrus (BA 21 {-57, -30, 2} to (BA 22) {50, -30, 4}, 21 voxels). Greater activation with alien-undistorted feedback was also observed in the following regions: bilateral inferior frontal gyri (BA 47), posterior cingulate gyrus (BA 31), primary occipital cortex, right hippocampus, thalamus, caudate, putamen, brainstem and cerebellum.

The self-undistorted condition was associated with greater activation in the right superior temporal gyrus (BA 22 {-57, -4, 4})

**Self-undistorted (yellow) v self-distorted (red)**

Superior temporal gyrus BA 42 {53, -13, 9}

![SSQ](image1)

**Self-undistorted (yellow) v alien-undistorted (red)**

Superior temporal gyrus BA 22 {-57, -17, 4}

![SSQ](image2)

**Figure 3.** Comparisons between feedback conditions revealed differential engagement of the lateral temporal cortices and other regions in the self-monitoring network. The self-distorted condition showed greater activity relative to the self-undistorted condition in the superior and middle temporal gyri bilaterally, while the self-undistorted condition had greater activity in the left temporal pole and smaller clusters in the lateral temporal gyr. The alien-undistorted condition also showed greater activation relative to self-undistorted feedback in the lateral temporal cortices bilaterally, while the self-undistorted condition was associated with greater activation in the right superior temporal gyrus relative to alien-undistorted feedback. Transverse images are shown in radiological convention on a standard template in Talairach and Tournoux space. Plots of the median SSQ ratio, which provides a measure of the BOLD response, are presented for each condition at specified voxels.

**Self-distorted versus Alien-undistorted Feedback.** The self-distorted condition revealed greater activity relative to alien-undistorted feedback in the lateral temporal cortices bilaterally: left superior temporal gyrus (BA 42 {-53, -17, 4}, 10 voxels), and right middle temporal gyrus (BA 21 {-57, -10, -7}, 10 voxels) and superior temporal gyrus (BA 42 {43, -20, 4} to {-57, -20, 15}, 84 voxels), as well as in the primary occipital cortex and cerebellum. In contrast, alien-undistorted feedback was associated with greater activity in the left middle temporal gyrus (BA 21 {-53, 0, 18}, 7 voxels; and {-53, -7, -7}, 17 voxels) and superior temporal gyrus (BA 42 {-57, -20, 20}, 27 voxels; and {-57, -30, 26}, 6 voxels), as well as in the right thalamus relative to self-distorted feedback. Full coordinates are available as supplementary material.

**Effects of Attribution Response**

To examine the neural correlates associated with the accuracy of attribution, within the self-distorted feedback condition, trials which subjects correctly identified the feedback (as self) were associated with greater activation in the right superior temporal gyrus (two clusters: BA 22 {53, 4, 4}, cluster size 5 voxels; and extending from BA 22 {-61, -26, -2} to (BA 42) {-57, -39, 15}, cluster size 47 voxels), left superior temporal gyrus (extending from BA 22 {-57, 17, 9} to BA 42 {-57, -37, 26}, cluster size 169 voxels), and occipital cortex (BA 18, extending from {0, -82, -7} to {4, -73, 4}, cluster size 46 voxels) relative to trials in which subjects made incorrect external misattributions (Fig. 4). There were no areas that were more activated in association with misattributions.

No difference in activations was found with the contrasts of trials in which subjects made correct attributions within the self-distorted and self-undistorted feedback conditions, nor within the self-distorted and alien-undistorted feedback conditions.

**Discussion**

The present study sought to examine the neural correlates of verbal self-monitoring using fMRI. Auditory verbal feedback was manipulated while subjects read aloud who were required to make a decision about the source of the speech they heard. The use of a clustered acquisition sequence allowed each trial of the task to be performed in relative silence, and the event-related design permitted a randomized presentation of the stimuli and individual analysis of the responses. Three main observations were evident from the fMRI data. First, as hypothesized,
a mismatch between expected and actual verbal feedback was associated with greater lateral temporal activation. Secondly, the mismatch engaged not only the temporal cortices but also other cortical and subcortical regions in the self-monitoring model (Frith, 1992). Thirdly, as each condition engaged the temporal cortices and other areas in the self-monitoring network, the distinction between conditions was derived from relative differences which may have reflected the processing of different features of the verbal feedback.

The first set of hypotheses was that the self-distorted and alien-undistorted feedback conditions would engage verbal self-monitoring and lead to increased activation in the lateral temporal cortices. In accordance with the earlier PET study (McGuire et al., 1996), both conditions elicited greater activation bilaterally in the lateral temporal cortices relative to self-undistorted feedback. The increases were localized to the middle and superior temporal gyri. In rhesus monkeys, the anterior portion of the superior temporal gyrus shows the greatest selectivity for monkey vocalizations (Rauschecker and Tian, 2000), and the same region is implicated in processing of human speech (Belin et al., 2000; Binder et al., 2000). Moreover, microelectrode recordings have identified neurons in the auditory cortex which respond specifically to externally-generated vocalizations (Muller-Pruess and Ploog, 1981). The engagement of these regions during verbal self-monitoring is consistent with its putative specialization for auditory speech processing.

Imagining another person speaking (McGuire et al., 1996) and monitoring externally generated speech (Démonet et al., 1992; Zatorre et al., 1992; Binder et al., 1995) have been associated with increased activation in the left temporal cortex. In particular, discerning the phonological or semantic features of speech elicits left temporal activation (Démonet et al., 1992; Zatorre et al., 1992; Binder et al., 1995; Jancke et al., 2002), while perceiving its prosodic features particularly engages the right temporal cortex (Ross and Mesulam, 1981; Zatorre et al., 1992; Mitchell et al., 2003). Furthermore, attending to auditory stimuli activates the temporal cortices bilaterally relative to processing the same stimuli but attending to a visual stimulus (Woodruff et al., 1996). The bilaterality of the temporal activation observed in the present study is consistent with the involvement of both of these features of auditory processing in verbal self-monitoring.

There was also greater engagement of subcortical regions, proposed components of the self-monitoring model (Frith, 1992). During the alien-undistorted condition, the thalamus showed increased activation relative to self-undistorted feedback. The thalamus has a key role in the auditory system with projections to extensive areas in the lateral temporal cortex (Middlebrooks and Zook, 1983; Cetas et al., 1999; Huang and Winer, 2000), and the pulvinar nucleus in particular projects to all aspects of the superior temporal gyrus (Burton and Jones, 1976; Veteran and Pandya, 1998). The inferior colliculus and medial geniculate body, essential components of the ascending auditory pathway, have reciprocal projections to other nuclei within the thalamus and to the lateral temporal cortices (LeDoux et al., 1987; Rauschecker and Tian, 2000; Winer et al., 2002). The thalamus has reciprocal connections with the cerebellum which has been implicated in speech generation and perception (Petersen et al., 1988; Fox et al., 1996; Desmond and Fiez, 1998), and the midline cerebellum also showed greater activity during the alien-undistorted condition.

The second set of hypotheses was that the self-distorted condition would place greater demands on verbal self-monitoring than the alien-undistorted feedback condition (Johns et al., 2001), and thus be associated with even greater activation in temporal cortices. In general, healthy volunteers reported that the self-distorted condition was more difficult, and their performance measures showed more erroneous responses and slower reaction times when they did make a correct response. Comparison of the self-distorted and alien-undistorted conditions revealed that the former was associated greater activation in voice processing regions in the superior temporal cortices bilaterally (Belin et al., 2000; Binder et al., 2000), particularly on the right side which may have reflected greater attention to the prosodic features of the self-distorted feedback (Ross and Mesulam, 1981; Zatorre et al., 1992; Mitchell et al., 2003). Conversely, the alien-undistorted condition was associated with greater engagement of the left temporal pole and a posterior portion of the left superior temporal gyrus. These regions are implicated in semantic knowledge and phonological processing (Kellenbach et al., 2005; Majerus et al., 2005). In particular, the left temporal pole is associated with taking the social perspective of a third person (Ruby and Decety, 2004) and, in monkeys, this region shows a species-specific response to the vocalizations of other monkeys (Poremba et al., 2004). The greater engagement of the left temporal pole with alien-undistorted feedback provides support for a self versus non-self/other distinction within the temporal cortices.

The alien-undistorted condition also showed greater activation in the hippocampus relative to the self-undistorted condition. In the PET study of this task (McGuire et al., 1996), a trend towards activation of the hippocampus was found during the early scans, suggesting that the task had particularly engaged self-monitoring processes when subjects initially encountered the new auditory feedback conditions which were presented consecutively within a block. In the present study, trials were presented randomly, and subjects could not anticipate the nature of the next feedback condition. The random presentation may have led to greater engagement of the hippocampus across all the conditions with insufficient differential activation that could be detected with the present study design.

The third set of hypotheses and contrasts focused on subjects’ behavioural performance during the self-monitoring conditions. Impaired self-monitoring has been proposed to be the neuropsychological basis of auditory verbal hallucinations in schizophrenia in which individuals may mis-identify their own inner speech as alien and thus perceive it as an external voice (Feinberg, 1978; Frith and Done, 1989; Frith, 1992). In verbal self-monitoring paradigms, patients with schizophrenia who are acutely psychotic are more likely to attribute the feedback to another person during the self-distorted condition relative to patients in remission from an acute psychotic episode and healthy individuals (Cahill et al., 1996; Johns and McGuire, 1999; Johns et al., 2001). In the present study, subjects showed the slowest reaction times for correct responses in the self-distorted feedback condition indicating the more difficult nature of this condition.

The event-related design allowed the cerebral activations associated with each individual trial to be classified according to whether the subject’s perception of the speech source was correct or incorrect. The inclusion of an ‘unsure’ option avoided subjects being asked to make a forced choice when they were
actually uncertain. Trials associated with unsure responses were excluded from this fMRI analysis, so the contrast was restricted to trials where the subject was confident that speech was either self- or externally generated. The correlates of misattribution errors during the self-distorted condition, when subjects misidentified their own speech as alien, were of particular interest. External misattributions during self-distorted feedback were associated with significantly reduced bilateral superior temporal activation relative to correct attributions. This finding was not wholly in support of our hypothesis. However, as the number of observations in this contrast was low, the power and reliability of the data from the present study were reduced, and this question requires further investigation. Interestingly, though, comparisons of correct attributions for each of the feedback conditions showed no significant differences between conditions. These observations suggest that correct attributions were associated with comparable activations in the lateral temporal cortices, irrespective of whether the attribution was to self or other.

There are a number of limitations in the present study. Although the conditions were presented in a factorial design (with two sources of feedback: self and alien; and two sources of distortion: none and with distortion) so that subjects would not be able to systematically recognize the source of the stimuli, in terms of self-monitoring processes, there was a single self condition (self-undistorted feedback) and three conditions with unusual feedback (self-distorted, alien-undistorted, and alien-distorted). The fMRI data analysis was designed to account for these differences, and the conditions were compared categorically. In each of the conditions, subjects produced an overt verbal response. A design with two sources of verbal generation [none (no response) and overt response] would have allowed a factorial analysis of the effects of self-generation. The effects of self-monitoring in the absence of overt verbal production has been investigated separately (Allen et al., 2004). It is possible that subjects could have obtained information about their true speech output via bone conduction. However, the feedback was instantaneous and of sufficient volume that one debriefing subjects said they could not perceive a bone conduction signal. As well, any effect would have applied similarly to all conditions. Another limitation was the level of pitch distortion which was determined from behavioural data (Johns et al., 2001) to elicit a sufficient number of misattribution errors for fMRI analysis. With a variety of pitch-distortion levels, a range of behavioural responses may have been observed. However, these additional conditions would necessarily extend the fMRI scans which is another consideration.

Perhaps though subjects were paying less attention during the unusual feedback conditions (self-distorted, alien-undistorted and alien-distorted) and had been merely attributing their feedback to ‘other’. It should be noted that a misattribution error for the ‘self’ conditions (self-undistorted and self-distorted) reflects a failure to recognize the source as oneself with an external misattribution to someone else, while a misattribution error for the ‘alien’ conditions (alien-undistorted and alien-distorted) is the converse, a failure to recognize the source as being from someone else and with a self-misattribution. If subjects were externally misattributing any feedback that appeared to be unusual, then the external misattribution (deciding that the feedback was ‘other’) during the self-distorted condition should be equivalent to the correct attributions made during the alien-distorted feedback condition, which is the most comparable experience, and perhaps also during the alien-undistorted feedback condition. Instead, subjects made half as many external misattribution errors during the self-distorted feedback condition as correct external attributions during the alien-distorted as well as alien-undistorted feedback conditions. The pattern of behavioural responses indicates that this is not merely attributable to poor attention with an externalizing response bias to ‘other’.

Furthermore, if the activation in the present experiment were simply a function of increased attention to an unusual auditory stimulus, a similar activation across the self-distorted, alien-undistorted and alien-distorted conditions would be expected. Instead, the pattern of activation in the temporal cortices associated with each condition relative to self-undistorted trials was different in each case, and there were additional differences when the self-distorted and alien-undistorted conditions were compared with each other.

Another limitation of the present study was the presence of scanner noise between trials. One disadvantage of fMRI relative to PET is that its image acquisition is associated with an acoustic noise that can make it difficult for subjects and investigators to hear overt verbal responses (Amaro et al., 2002). A clustered acquisition sequence was used which incorporated brief periods of silence (Eden et al., 1999; Hall et al., 1999) which ensured that subjects were able to make overt articulations and hear their responses in the absence of noise, as well as reducing the risk of artefacts secondary to articulation-related head movement (Fu et al., 2002). In addition, all the subjects in the present study were male volunteers, which may limit the generalizability of the findings.

In summary, these data suggest that verbal self-monitoring involves a network of areas implicated in the processing of auditory verbal material, in particular the lateral temporal cortices. While this network appears to be engaged independently of whether the monitoring is successful or unsuccessful, the correct recognition of self-generated speech seems to be associated with greater activation in the temporal cortices.

Notes

This work was supported by a Wellcome Trust Travelling Fellowship to C.H.Y.F. We would also like to thank the radiographers and Dave Gasston at the MRI Centre, Maudsley and South London NHS Trust, for their expert assistance; Drs Felix Beacher and Matthew Broome for their assistance with the word stimuli; and Dr Simon Meara for providing the ‘alien voice’.

Address correspondence to Dr C. Fu, Institute of Psychiatry, Division of Psychological Medicine and Social, Genetic and Developmental Psychiatry Centre, 103 Denmark Hill, London SE5 8AF, UK. Email: c.fu@iop.kcl.ac.uk.

References


